# Proceedings
# NeuroIS Retreat 2018

Vienna, Austria | June 19-21 | www.NeuroIS.org

Fred Davis, René Riedl, Jan vom Brocke,
Pierre-Majorique Léger, Adriane Randolph (Eds.)

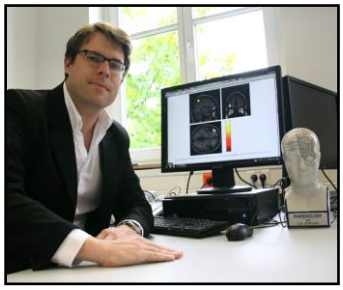
Proceedings of the NeuroIS Retreat 2018: Abstracts.
The final proceedings will be published by Springer.

## Bernd Weber – Keynote

# Translational Behavioral Neuroscience: The Use of Neuroscientific Insights to Improve Public Welfare

Over the last decades, neuroscience has provided fundamental insights into the processes underlying the development and heterogeneity of human behavior. With increasing knowledge about the neurobiological processes and computations, scholars have begun to investigate possible applications of neuroscientific methods and insights in many different domains. Within this talk I want to provide examples and discuss the usefulness of these insights for improving environments in a way supporting human development and decision making. Core examples will relate to nutrition and food choice, decisions under risk and uncertainty or prediction of behavior (change).

*Prof. Dr. Bernd Weber has a degree in medicine and he completed his habilitation in experimental neurology. He is the Director of the Center for Economics and Neuroscience, University of Bonn, Germany. He has published over 150 peer-reviewed journal articles, and his research appeared in highly prestigious journals such as Science, Nature, Neuron, Brain, and PNAS. Also, he is a former co-editor of the Journal of Neuroscience, Psychology, and Economics. Dr. Weber was one of the neuroscience experts who participated at the inaugural NeuroIS Retreat in 2009, and he has been actively contributing to the development of the NeuroIS field in the past decade.*



# Christian Montag – Hot Topic Talk

## The neuroscience of smartphone / social media usage and the growing need to include methods from 'Psychoinformatics'

The present work gives a brief overview of the current state of affairs in the investigation of the neuroscientific underpinnings of social media use. Such an overview is of importance because individuals spend significant amounts of time on these 'social' online channels. Despite several positive aspects of social media use, such as the ability to easily communicate with others across long distances, it is clear that detrimental effects on our brains and minds are possible. Given that much of the neuroscientific and psychological research conducted up to now relies solely on self-report measures to assess social media usage, it is argued that neuroscientists/psychologists need to include more digital traces resulting from human-machine/computer interaction, and/or information shared by individuals on social media, in their scientific analyses.

*Prof. Dr. Christian Montag is Professor for Molecular Psychology at Ulm University in Ulm/ Germany and Visiting Professor at the University of Electronic Science and Technology of China in Chengdu/China. He studied psychology in Giessen/Germany and then worked on his PhD and habilitation dissertation in Bonn/Germany.*

*Christian Montag is interested in the molecular genetics of personality and emotions. He combines molecular genetics with brain imaging techniques such as structural/functional MRI to better understand individual differences in human nature. Adding to this he conducts research in the fields of Neuroeconomics and addiction including new approaches from Psychoinformatics.*

*Currently Christian Montag is on the editorial board of Addictive Behaviors (Elsevier), Personality Neuroscience (Cambridge University Press) and International Journal of Environmental Research and Public Health (MDPI). He is co-editor of the book series Studies in Neuroscience, Psychology and Behavioral Economics. He is (co-)author of more than 150 peer-reviewed research articles including works in Science, Neuropsychopharmacology and NeuroImage. Moreover, he is author of several popular books. His last popular book is called Homo Digitalis.*

# NeuroIS: A Survey on the Status of the Field

Thomas Fischer[1], Fred D. Davis[2], René Riedl[1, 3]

[1] University of Applied Sciences Upper Austria, Steyr, Austria
{thomas.fischer, rene.riedl}@fh-steyr.at
[2] Texas Tech University, Lubbock, USA
fred.davis@ttu.edu
[3] University of Linz, Linz, Austria

**Abstract.** NeuroIS has emerged as a research field in the Information Systems (IS) discipline over the past decade. Since the inaugural NeuroIS Retreat in 2009, 166 individuals participated at this annual academic conference to discuss research and development projects at the nexus of IS and neuroscience research. Motivated by the fact that the NeuroIS Retreat celebrates its 10-year anniversary in 2018, we invited all 166 former participants of the NeuroIS Retreat to state their opinions in an online survey on the development of the field and its future. In this paper, we summarize the answers of N = 60 respondents regarding NeuroIS topics and methods.

**Keywords:** Brain · Methods · NeuroIS · Status · Survey · Tools · Topics

## 1 Introduction

The first NeuroIS Retreat took place in Gmunden, Austria, in 2009. Since then, the NeuroIS community has grown and in 2018 this annual academic conference celebrates the 10-year anniversary in Vienna, Austria. A total of 166 individuals attended this forum for the presentation and discussion of research and development projects in the last decade, and thereby contributed to the development of the field. Motivated by the fact that the NeuroIS Retreat exists for 10 years now, we developed an online survey to ask all former conference participants about their perspectives on the status of the field. In this paper, we present major results of this survey related to NeuroIS topics and methods. Specifically, we investigated the participants' perspectives on topics and methods thatare currently studied and applied, and what they think about future topics and methods.

## 2 Survey Characteristics and Sample Demographics

Using the online survey tool SoSci Survey, we conducted a survey amongst a population of all 166 previous participants of the NeuroIS Retreat 2009-2017 in the period

12/04/2017-02/06/2018. The survey contained questions related to impressions of the past developments in the field, but also gave respondents the opportunity to report on their future NeuroIS research and their expectations for the field. Overall, it took about ten minutes to complete the survey. We were able to gather 60 complete responses, amongst 152 individuals who are still involved in academic research (response rate of 39.5%). The remaining 14 individuals are not active researchers anymore and it was not possible to contact them in the context of this study.

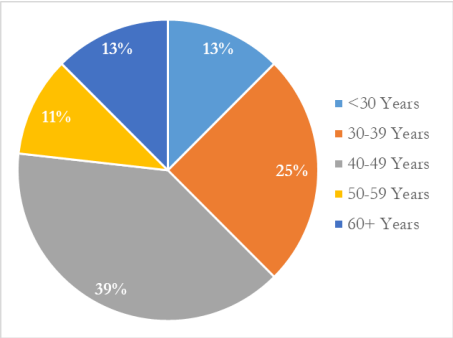



**Fig. 1.** Share of respondents per age group (N = 60)

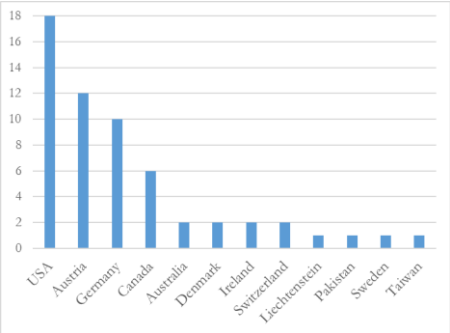



**Fig. 2.** Number of respondents per country of employment (N = 60)

Amongst the respondents, 75% were male and a majority of 64% were between 30 and 49 years old (see Figure 1). We also asked respondents to indicate the country were they are currently employed (see Figure 2). The results show that most respondents are currently either employed in German-speaking countries (25 individuals are from Austria, Germany, Switzerland, and Liechtenstein) or North America (24 individuals are from the USA and Canada).

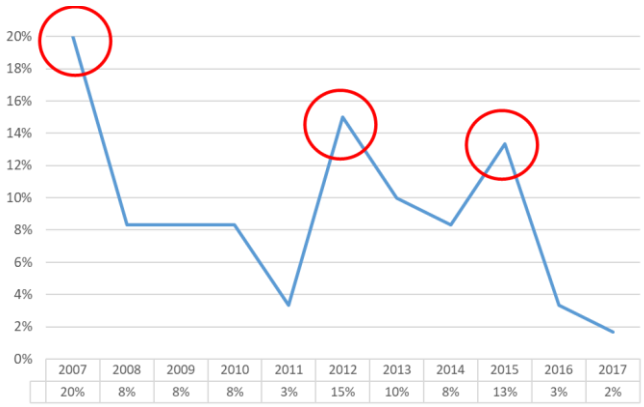



**Fig. 3.** Share of respondents per year in which they first came into contact with NeuroIS (N = 60)

We also wanted to know the current academic position of our respondents, which revealed that 39% were full professors, followed by 19% who were PhD candidates and

17% each who were either associate professor or assistant professor. This finding indicates that the field is not only interesting to a selected group of established researchers, but also allows new researchers to enter the arena, such as early-stage researchers. This finding is substantiated by the fact that only 20% of our respondents have been affiliated with NeuroIS since its first appearance in 2007; there is a substantial number of researchers who entered the field later (e.g., 2012 and 2015, see Figure 3).

Most of these individuals (85%) came into touch with NeuroIS through personal contacts (e.g., PhD students through their professors who had previously attended the NeuroIS Retreat), but also NeuroIS publications were an important source of information (28%). The website *www.NeuroIS.org* and conference calls were also of some importance (point of contact for 13% of respondents each), but not comparable to word-of-mouth spread throughout the NeuroIS community and related communities such as the more general IS community.

## 3 Topics

We asked respondents about the NeuroIS topics on which they had focused in their previous research and the topics they think were most important in NeuroIS research in the past decade. As our respondents had the possibility to indicate more than one topic (or construct), we ended up with a list of more than 40 different NeuroIS topics. Here we report the topics which were mentioned by at least 10% of our respondents as a current or future focus in their research or as being amongst the most important NeuroIS topics in the past decade. Through some abstractions (e.g., grouping "emotional responses" and "affective processing" into the category "Emotions"), we ended up with eight main topics (see Figure 4).

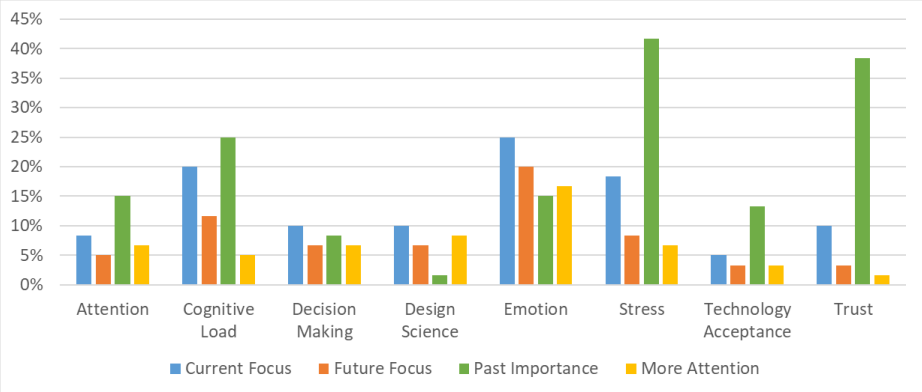



**Fig. 4.** NeuroIS topics with share of respondents who currently focus on them (blue bar) and will focus on them in the future (orange bar), importance of the topic in the past decade (green bar) and calls for more attention in future research (yellow bar) (N = 60)

We first looked at the current and future focus in the research of our respondents (see the blue and orange bars in Figure 4) and found that topics which are established

in neuroscience (or related fields such as neuropsychology or neuroergonomics) such as cognitive load, emotions, and stress, are also amongst the most popular topics in NeuroIS research. As shown in Figure 4, it can be expected that there will be a stronger focus on emotion in future research. In the case of other popular topics (e.g., technology acceptance or trust), our respondents were not so certain whether they will still focus on these topics in their future research. These findings are also in line with a recent review, which showed that cognitive and emotional processes have been the main focus in the extant NeuroIS literature, while decision-making processes and social processes were of lower importance [1].

In addition, we asked the respondents to indicate the topics that they felt had been the most important ones in the first decade of NeuroIS research (green bars, Figure 4) and whether these topics should receive more attention (yellow bars, Figure 4). Interestingly, emotion is not amongst the top 3 of the most important topics. Instead, most respondents felt that trust was amongst the most important topics, in addition to stress and cognitive load. This result is plausible because early NeuroIS publications in top IS journals had a focus on trust, such as [2]. Still, emotion as a topic received the most votes (i.e., 17%) when it came to the topics that should receive more attention in future research.

We further analyzed the topics that respondents focus on in their current and will focus on in their future research, based on two respondent characteristics, namely their tenure status and their country of employment, grouped into continents. For the tenure status, we looked at the differences between full professors (39% of our respondents) and the remaining respondents. For the country of employment, we looked at differences between researchers from Europe (50% of our respondents) and North America (40% of our respondents).

Regarding the current and future research topics of full professors, we found noteworthy differences (see Figure 5). In general, most respondents who are currently not full professors are uncertain about the topics on which they will focus in their future research (which can, for example, be explained by the uncertainty of the future funding of their research). Full professors rather than the remaining respondents indicated that they will focus more strongly on emotions, as well as stress, in their future research, while decision-making and trust are topics of lower interest. For most other topics (e.g., attention, cognitive load, or design science) we observe equal interest by full professors in the future.

We also found differences regarding the thematic focus of researchers from Europe and North America (Figure 6). While emotions and stress are more prevalent topics for European researchers, particularly design science is a topic that is more prevalent in the research of American researchers (note that design science, as defined in our research context, does not necessarily imply systems engineering activities, which are often typical for researchers from German-speaking countries, [3]). There will also likely be some shifts in the thematic focus, with European researchers focusing more strongly on attention and cognitive load research in the future, while American researchers will likely more strongly focus on decision-making.

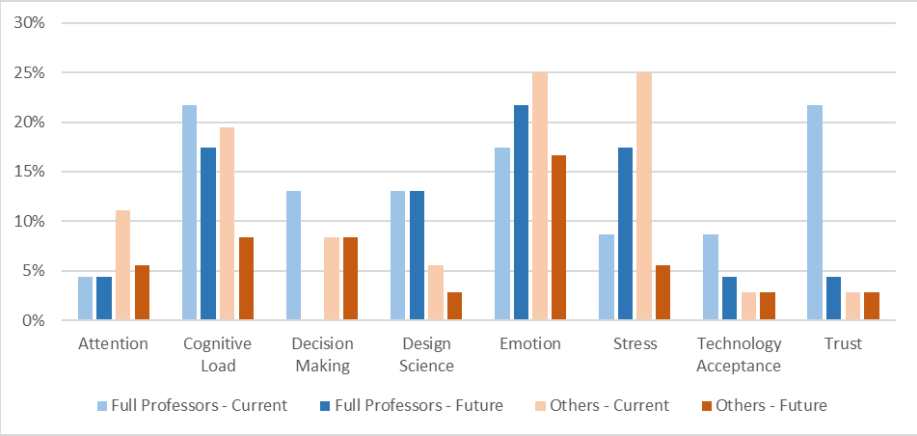



**Fig. 5.** NeuroIS topics with share of full professors who currently focus on them (light blue bar) and will focus on them in the future (dark blue bar), and researchers with a different tenure status who currently focus on them (light red bar) and will focus on them in the future (dark red bar) (N = 59)

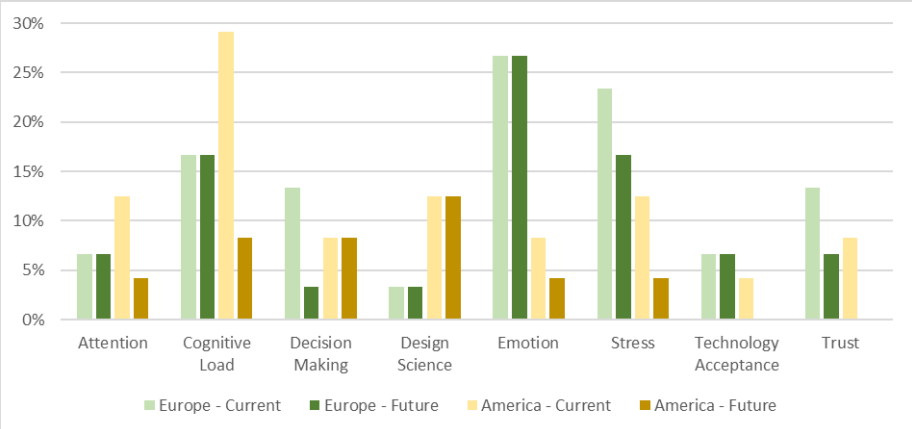



**Fig. 6.** NeuroIS topics with share of respondents from Europe who currently focus on them (light green bar) and will focus on them in the future (dark green bar), and researchers from North America who currently focus on them (light yellow bar) and will focus on them in the future (dark yellow bar) (N = 54)

## 4 Methods

We also asked the respondents about data collection methods they had previously used in their NeuroIS research, which methods they may use in the future, and whether they thought that certain methods should receive more or less attention in future NeuroIS research. We included a total of 13 data collection methods in our survey (i.e., blood

pressure, heart rate related-measures, eyetracking, EMG, EEG, fMRI, NIRS, skin conductance-related measures, hormone measures based on blood, urine, or saliva, neurological patients, and transcranial direct current stimulation). In Figures 7 and 8, we have summarized the results for each of these methods regarding (1) how many respondents have used them before ("previous use", blue bar), and (2) how those respondents who did not use a method before, intend to use it in the future ("intended use", orange bar; e.g., 20% of the respondents used hormone measures from saliva before and an additional 20% intend to use it in the future). In the case of previously used methods, eyetracking is on top with 58% of respondents indicating that they had used this method in their research. For intended use, hormone measurements based on saliva are in the lead, with 20% of respondents indicating that they would like to use this method in their future research (see Figure 7).

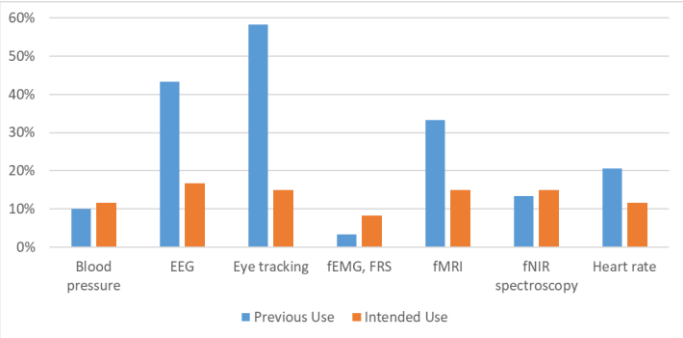



**Fig. 7.** NeuroIS methods with share of respondents who have previously used them (blue bar) and intent to use them in the future amongst previous non-users (orange bar) (N = 60), Part 1

In addition to eyetracking, which is widely employed and will also likely receive further attention in the future, particularly measures that collect data related to processes of the central nervous system (i.e., EEG, fMRI, NIRS, tDCS, and to some extent neurological patients) will be part of the future research of our respondents. It is interesting to see though, that saliva measurements may become more popular in the future as they can, for example, be used to measure physiological stress based on alpha amylase levels as indicator (e.g., [4]). Because the use of saliva samples, if compared to central nervous system measurements, implies less research effort and causes lower costs, it seems that many NeuroIS researchers base their research tool selection on pragmatic reasons. Why the intended use of measurements related to autonomic nervous system activity (e.g., heart rate, skin conductance) is rather low in our sample (despite its enormous potential in IS research, see [5]) remains an open question that deserves further investigation.

Some respondents also mentioned additional methods that should be of importance in future NeuroIS research including voxel-based morphometry (VBM), Magnetoencephalography (MEG), genetic measures, measurements made using data from everyday devices (e.g., smartwatches, see [6]), combinations of methods (e.g., eyetracking and fMRI, see [7]) and behavioral measures such as mouse cursor tracking.

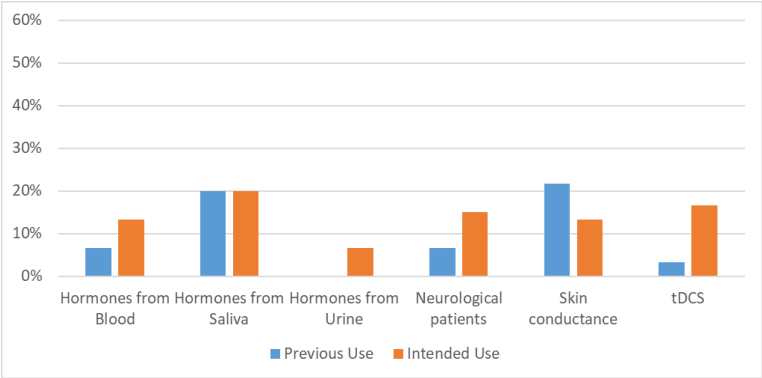
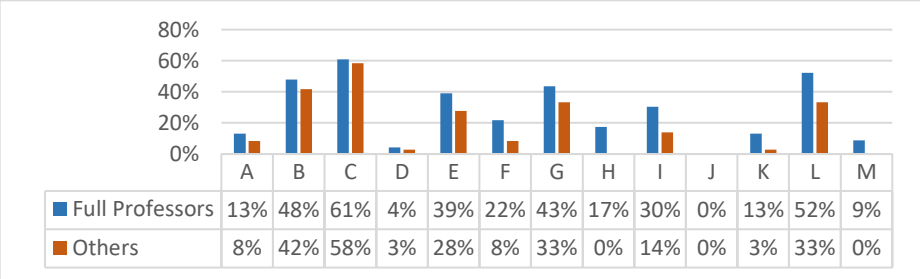



**Fig. 8.** NeuroIS methods with share of respondents who have previously used them and intend to use them in the future amongst previous non-users (N = 60), Part 2

In Figures 9 and 10, we provide an overview of the differences concerning the previous use of NeuroIS methods among our respondents based on tenure status (Figure 9) and country of employment (Figure 10).




| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Full Professors | 13% | 48% | 61% | 4% | 39% | 22% | 43% | 17% | 30% | 0% | 13% | 52% | 9% |
| Others | 8% | 42% | 58% | 3% | 28% | 8% | 33% | 0% | 14% | 0% | 3% | 33% | 0% |

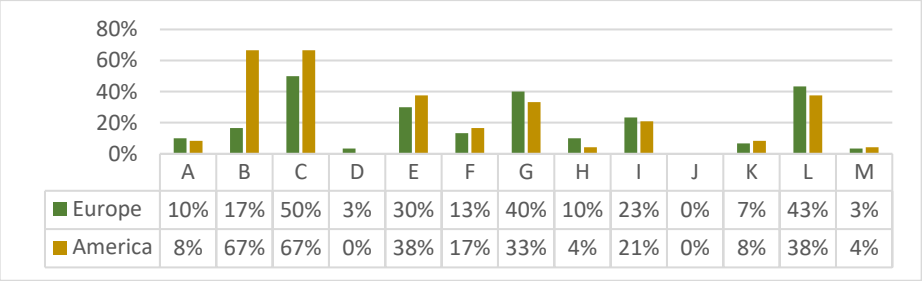
**Fig. 9.** NeuroIS methods with share of respondents with full professor status who have previously used them (dark blue bar) and share of respondents with other tenure status who previously used them (dark orange bar) (N = 59)

Legend: (A) blood pressure, (B) EEG, (C) Eyetracking, (D) fEMG, (E) fMRI, (F) fNIRS, (G) HR, (H) Hormones from Blood, (I) Hormones from Saliva, (J) Hormones from Urine, (K) Neurological Patients, (L) Skin Conductance, (M) tDCS

Based on tenure status, we hardly find differences, though the share of full professors who have used fNIRS (F), hormones from blood (H) or saliva (I), as well as neurological patients (K), in their research is considerably larger than the share of respondents with other tenure status. This could, for example, be explained by the complexity of getting access to the involved materials and data (e.g., in the case of hormones and neurological patients) or the novelty and cost of research methods (e.g., fNIRS), which makes access to these methods harder for individuals with lower tenure status.

We also analyzed differences in previous method use based on the country of employment, again clustered by continent (Figure 10). We find tendencies for European researchers to more frequently employ methods that can be used to measure the activity

of the autonomic nervous system (e.g., (G) heart rate or (L) skin conductance), while North American researchers more frequently employ methods that can be used to measure the activity of the central nervous system (e.g., (B) EEG or (E) fMRI). The largest differences can be found for (B) EEG and (C) Eyetracking, which are more frequently used by North American researchers. Future research may determine the reasons for the observed differences.




| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Europe | 10% | 17% | 50% | 3% | 30% | 13% | 40% | 10% | 23% | 0% | 7% | 43% | 3% |
| America | 8% | 67% | 67% | 0% | 38% | 17% | 33% | 4% | 21% | 0% | 8% | 38% | 4% |

**Fig. 10.** NeuroIS methods with share of respondents from Europe who have previously used them (dark green bar) and share of respondents from North America who previously used them (dark yellow bar) (N = 54)

Legend: (A) blood pressure, (B) EEG, (C) Eyetracking, (D) fEMG, (E) fMRI, (F) fNIRS, (G) HR, (H) Hormones from Blood, (I) Hormones from Saliva, (J) Hormones from Urine, (K) Neurological Patients, (L) Skin Conductance, (M) tDCS

## 5 Conclusion

Based on an online survey among N = 60 former participants of the NeuroIS Retreat, we found that emotional processes will likely be a key topic, eventually *the* key topic, in future NeuroIS research. Methodologically, we found that eyetracking measures and brain-related measures such as EEG or fMRI will be of high relevance in the future. These findings are in line with observations in the NeuroIS literature, as emotional processes have been of major importance in previous NeuroIS research [1], and NeuroIS publications in the most prestigious IS journals have often applied brain-related measures such as fMRI (e.g., [2, 8–12]). Importantly, our survey also revealed the interest of NeuroIS researchers in methods which have not been used frequently thus far, such as EMG and FRS (Face Recognition Software, e.g. to determine user emotion, for details see [13]). It seems that NeuroIS researchers have realized that these and further methods are well suited to reveal insights into the NeuroIS topics of the future (e.g., [14]).

**Acknowledgements.** This research was funded by the Upper Austrian government as part of the PhD program “Digital Business International”, a joint initiative between the University of Applied Sciences Upper Austria and the University of Linz.

## References


1. Riedl, R., Fischer, T., Léger, P.-M.: A Decade of NeuroIS Research: Status Quo, Challenges, and Future Directions. In: Kim, Y.J., Agarwal, R., Lee, J.K. (eds.) Proceedings of the Thirty Eighth International Conference on Information Systems (2017)
2. Dimoka, A.: What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. MIS Quarterly 34, 373–396 (2010)
3. Heinrich, L.J., Riedl, R.: Understanding the dominance and advocacy of the design-oriented research approach in the business informatics community: A history-based examination. Journal of Information Technology 28, 34–49 (2013)
4. Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., Grover, V.: NeuroIS - Alternative or Complement to Existing Methods? Illustrating the Holistic Effects of Neuroscience and Self-Reported Data in the Context of Technostress Research. Journal of the Association for Information Systems 15, 723–753 (2014)
5. Riedl, R., Davis, F.D., Hevner, A.: Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement. Journal of the Association for Information Systems 15, i–xxxv (2014)
6. Fischer, T., Riedl, R.: Lifelogging as a viable data source for NeurolS researchers: A review of neurophysiological data types collected in the lifelogging literature. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., Randolph, A.B. (eds.) Information Systems and Neuroscience. Gmunden Retreat on Neurois 2016, 165–174. Springer-Verlag New York Inc (2016)
7. Vance, A., Jenkins, J.L., Anderson, B.B., Bjornn, D.K., Kirwan, C.B.: Tuning out Security Warnings: A Longitudinal Examination of Habituation through fMRI, Eye Tracking, and Field Experiments forthcoming (2018)
8. Riedl, R., Hubert, M., Kenning, P.H.: Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers. MIS Quarterly 34, 397–428 (2010)
9. Anderson, B.B., Vance, A., Kirwan, C.B., Jenkins, J.L., Eargle, D.: From Warning to Wallpaper: Why the Brain Habituates to Security Warnings and What Can Be Done About It. Journal of Management Information Systems 33, 713–743 (2016)
10. Jenkins, J.L., Anderson, B.B., Vance, A., Kirwan, C.B., Eargle, D.: More Harm Than Good? How Messages That Interrupt Can Make Us Vulnerable. Information Systems Research 27, 880–896 (2016)
11. Riedl, R., Mohr, P.N.C., Kenning, P.H., Davis, F.D., Heekeren, H.R.: Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. Journal of Management Information Systems 30, 83–114 (2014)
12. Warkentin, M., Walden, E.A., Johnston, A.C., Straub, D.W.: Neural Correlates of Protection Motivation for Secure IT Behaviors: An fMRI Examination. Journal of the Association for Information Systems 17, 194–215 (2016)
13. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS: Information Systems and the Brain. Springer, Berlin, Heidelberg (2016)
14. Koller, M., Walla, P.: Measuring Affective Information Processing in Information Systems and Consumer Research – Introducing Startle Reflex Modulation. In: George, J.F. (ed.) Proceedings of the 33rd International Conference on Information Systems (2012)

# Improving Security Behavior through Better Security Message Comprehension: fMRI and Eye-tracking Insights

Anthony Vance[1*], Jeffrey L. Jenkins[2], Bonnie Brinton Anderson[2], C. Brock Kirwan[2], and Daniel Bjornn[2]

[1]Temple University, Philadelphia, Pennsylvania, USA
[1]anthony@vance.name
[2]Brigham Young University, Provo, Utah, USA
[2]{jeffrey_jenkins, bonnie_anderson, kirwan, dbjornn}@byu.edu

**Abstract.** Security warnings are critical to help users make contextual security decisions. Unfortunately, users find these warnings hard to understand, and they routinely expose themselves to unintended risks as a result. Although it is straightforward to determine when users fail to understand a warning, it is more difficult to pinpoint *why* this happens. The goal of this research is to use eye tracking and fMRI to step through the building blocks of comprehension—attention, semantics, syntax, and pragmatics—for SSL and other common security warnings. Through this process, we will identify ways to design security warnings to be more easily understood.

**Keywords:** NeuroIS · eye-tracking · fMRI · comprehension · security messages.

## 1 Introduction

Users routinely disregard protective messages such as software security warnings [2; 3]. One reason for the ineffectiveness of warnings is the mismatch between security concerns and security behavior. For example, individuals' stated security concerns have been found to be inconsistent with their subsequent behavior in response to security warnings [11]. These empirical results confirm those of Crossler et al. [5], who called for research that explains the discrepancy between security intentions and behaviors.

One important factor contributing to the disconnect between security concerns and actual behavior is the lack of comprehension. For example, in the case of security warnings, although users may intend to behave securely, they may not comprehend a security warning, which may in turn lead them to make a choice that unintentionally exposes themselves to security risks.

Past research on comprehension of security warnings has highlighted the difficulty users have in understanding security warnings. Felt et al. [6] tested several iterations of text and design for SSL warnings in Google Chrome. They found that users routinely had difficulty determining the threat source and data risk, even after designing interventions to improve comprehension.

However, comprehension is not a binary event, but rather involves interrelated stages that lead to understanding. These stages include [9]:

1-Attention—focused mental resources on a certain object.

2-Semantics—the meaning of individual words and simple phrases.

3-Syntax—the structure of sentences that creates relationships between words.

4-Pragmatics—the application of past experience and knowledge to infer meaning.

The research objectives of this study are to: (1) use eye tracking, fMRI, and users' behavioral responses, through a series of complementary experiments, to determine failures at each of the above stages of comprehension for security warnings. Through this process, we will (2) identify ways to design security warnings to improve comprehension at each stage.

## 2 Planned Research and Expected Outcomes

### 2.1 Past Research on Comprehension of Security Warnings

Poor comprehension of security warnings is a common finding in the human–computer interaction literature. For example, researchers found that Android users paid attention to app permissions during installation only 17% of the time, and only 3% of users could correctly answer comprehension questions about permissions they saw [7]. Similarly, in a later study they found that users comprehended the threat source of SSL warnings in Chrome only 37.7% of the time, and comprehended even less what data was at risk. By changing the warning design based on recommendations from warning literature, they improved threat source comprehension nearly 12%. However, the design was not able to improve the comprehension of the risk to data [6].

We build on this past literature by applying behavioral information security to better understand and improve users' security behaviors [1]. Based on our findings, we expect to be able to determine more precisely where and why warning comprehension breaks down both from a neural and behavioral perspective. This will, in turn, allow us to create guidelines to improve comprehension in security warnings.

Previous work on comprehension using eye tracking found that more complex sentence structures result in poorer comprehension. For example, participants who read sentences with confusing ("ambiguous") syntax had poorer ability to answer simple questions about the sentences correctly compared to similar sentences that were changed slightly to be less confusing ("disambiguated"). Specifically, comprehension accuracy decreased by 15–38% when syntax was complex. This impaired comprehension was paired with significantly more re-reading of the ambiguous sentences (27–60% more time spent re-reading). In summary, not only does complex syntax impair comprehension, but re-reading is a reliable indicator of this impairment [4].

### 2.2 Description of Project and Expected Outcomes

To achieve our research objectives, we will record eye tracking data to step through the stages of comprehension (see Figure 1). Comparable to code debugging, we will

work through the different stages of comprehension to determine where comprehension is impeded. We will then improve warning designs to increase attention, ensure clear semantic and syntactic understanding, and promote pragmatic cognition. For example, at the level of attention, use of symbols or animation may help to improve overall attention. Similarly, semantic understanding may be improved through use of more familiar terms, or increased word frequency. By examining each stage individually, we expect to improve comprehension overall.

Eye tracking is an ideal tool for measuring the moment-by-moment allocation of attention. It is also used in psychology and linguistics to explore how people understand written language and to measure comprehension difficulty. For instance, words that are less familiar or unexpected (semantics) are looked at longer, and complex or confusing sentences (syntax) are re-read more often than are simple sentences [10]. In contrast to eye tracking, fMRI can provide information about the underlying neural and cognitive operations in attentional, semantic, and syntactic processing [8].

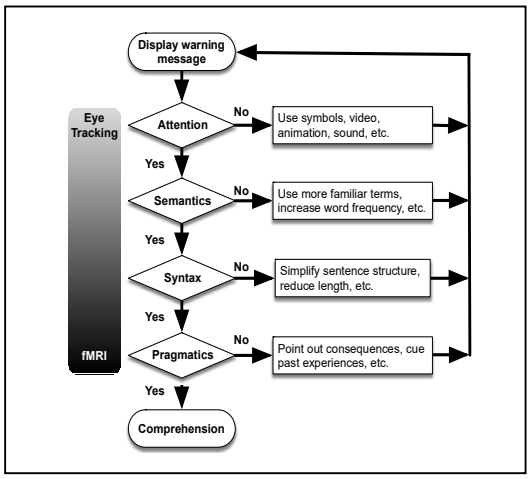



**Fig. 1.** Evaluating warnings at different stages of comprehension using eye tracking and fMRI.

### 2.3 Hypotheses

We propose an eye-tracking experiment that examines the influence of syntax on users' comprehension of warnings. We will examine whether changing the syntax of the warning to place the focus on different aspects of a data security breach. In addition to the usual focus on the attacker or the target website, we will also include a condition where the syntax of the warning shifts the focus to the consequences of ignoring the warning. We hypothesize that:

> Hypothesis 1 – changing the focus of the warning will result in significant differences in comprehension as evidenced by a significant difference in the number of regressions (i.e., rereading) across warning focus.

Eye movement regressions are often used as a non-conscious measure of reading comprehension. As such, they may be more sensitive to subtle differences in compre-

hension between the different warning focus conditions in our experiment. Additionally, syntax changes should result in differences in overt comprehension as measured by performance on post-hoc comprehension questions. We hypothesize that:

> Hypothesis 2 – eye regressions in turn will significantly predicted whether participants correctly understand the warnings as measured by performance on a post hoc comprehension quiz

# 3 Eye Tracking Pilot Study

## 3.1 Participants and Stimuli

A total of 43 college-age individuals (14 male, 29 female) participated in the study. Five participants were not able to participate because an accurate calibration was not obtained. Removing these five participants left the sample with 38 individuals (14 male, 24 female). Participants were given course credit for participating in the study.

Warnings were created by sampling four warning types from the Google Chrome browser and the Apple Safari browser, namely malware, phishing, SSL, and unwanted software. The text for the warnings was then manipulated by changing the subject, verb, and object of the statement. For example, warnings from Chrome focus on the attacker as the subject of the statement. An example of this focus can be seen from the SSL warning text, "Attackers might be trying to steal your information from expired.badssl.com (for example, passwords, messages, or credit cards)." Warnings from Safari focus on the website as the subject of the statement. Chrome warning text was manipulated to change the focus to the website and Safari warning text was manipulated to change the focus to the attacker.

Along with the focus on the attacker and the website, a third text condition focused on the potential consequences of ignoring the warning. For example, the chrome SSL warning could be changed to, "Your information from expired.badssl.com (for example, passwords, messages, or credit cards) might be stolen if you visit it." Text from Chrome and Safari warnings were manipulated to fit this design.

The four warning types (i.e., malware, phishing, SSL, and unwanted software) for two browsers (i.e., Chrome and Safari) across three different conditions (i.e., attacker focus, consequence focus, and site focus) provided 24 different warnings. All references to a specific website were changed to "this website" for ease of presentation.

The warning text was overlaid onto a mock warning image for each trial. Warning titles were created from the standard text from the warning type for each browser (e.g., "Your connection is not private" for the Chrome SSL warning and "This Connection Is Not Private" for the Safari SSL warning).

## 3.2 Task

Participants viewed each warning one at a time on the computer screen and then answered a question. Each trial began with a drift check, which required participants to

look at a circle on the top left part of the screen and press the spacebar to continue. The warning was then presented and participants read the warning and pressed the spacebar when they were ready to continue. The last part of each trial was the comprehension question which asked, "If this were a real threat and I ignored this warning, an attacker could," and then presented four answer options. Each of the answer options corresponded to a warning type:

- Phishing – "Trick me into installing malicious software or disclose personal information"
- Malware – "Install a dangerous program on my computer that could steal my information or delete my data"
- SSL – "See anything I send or receive from the website"
- Unwanted Software – "Install software that displays ads on my computer or make changes to my browser"

The answer options were presented in a random order for each trial. The full task consisted of 24 trials. The eye tracker was calibrated before the start of the task and after every 6 trials. Warnings were presented in a random order for each participant.

### 3.3 Planned Analysis

The results of this experiment will be analyzed by examining the behavioral and eye tracking measures of comprehension. For the behavioral analyses, we will calculate the proportions correct for warning type and warning focus separately. Repeated measures ANOVA tests will be run to test these factors individually in order to ensure a large enough number of trials for each bin.

For hypothesis 1, we will test whether warning focus predicts the number of regressions (i.e., rereading the text) by entering the total number of regressions for each trial as a dependent variable into a linear regression model with an independent variable of warning focus. For hypothesis 2, we will test whether the number of regressions predict accuracy on the comprehension test by entering trial accuracy as a dependent variable into a linear regression model with an independent variable of the total number of regressions in the trial. We will also use comprehension of other, non-security, messages and warnings for a comparison. Our post-study survey will contain that standard demographic, education and computer experience, as well as security risk questions, big five personality traits, and general risk propensity profile.

## 4 Conclusion

Users often respond inappropriately to security warnings. A significant factor in this failure is users' difficulty in comprehending warnings. The insights expected to be gained from this research have the potential to inform the design and evaluation of warnings that more effectively help users to respond to security threats, enhancing the information security of individuals and organizations.

## References


1 B.B. Anderson, C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard and A. Vance 2015. How polymorphic warnings reduce habituation in the brain: Insights from an fmri study. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* ACM, Seoul, South Korea.

2 B.B. Anderson, A. Vance, C.B. Kirwan, D. Eargle and J.L. Jenkins. How users perceive and respond to security messages: A neurois research agenda and empirical study. *European Journal of Information Systems 25*, 4 (2016), 364-390.

3 C. Bravo-Lillo, S. Komanduri, L.F. Cranor, R.W. Reeder, M. Sleeper, J. Downs and S. Schechter 2013. Your attention please: Designing security-decision uis to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* ACM, Newcastle, United Kingdom, 1-12.

4 K. Christianson, S.G. Luke, E.K. Hussey and K.L. Wochna. Why reread? Evidence from garden-path and local coherence structures. . *Quarterly Journal of Experimental Psychology 70*, 7 (2017), 1380-1405.

5 R.E. Crossler, A.C. Johnston, P.B. Lowry, Q. Hu, M. Warkentin and R. Baskerville. Future directions for behavioral information security research. *Computers & Security 32*, 1 (2013), 90-101.

6 A.P. Felt, A. Ainslie, R.W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris and J. Grimes. Improving ssl warnings: Comprehension and adherence. In *Proceedings of the Conference on Human Factors in Computing Systems* (2015).

7 A.P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin and D. Wagner 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* ACM, 3:1-3:14.

8 T.A. Keller, P.A. Carpenter and M.A. Just. The neural bases of sentence comprehension: A fmri examination of syntactic and lexical processing. *Cerebral Cortex 11*, 3 (2001), 223–237.

9 K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*, 3 (1998), 372-422.

10 K. Rayner. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology 62*, 8 (2009), 1457-1506.

11 A. Vance, B.B. Anderson, C.B. Kirwan and D. Eargle. Using measures of risk perception to predict information security behavior: Insights from electroencephalography (eeg). *Journal of the Association for Information Systems 15*, 10 (2014), 679-722.

# Neural Activity Related to Information Security Decision Making: Effects of Who is Rewarded and When the Reward is Received

Bridget Kirby[1], Kaitlyn Malley[1], and Robert West[1]
Department of Psychology and Neuroscience[1], DePauw University, Greencastle, USA
{bridgetkirby_2020, kaitlynmalley_2019, robertwest}@depauw.edu

**Abstract.** Breaches of information security resulting from cybercrime represents a significant threat to the security and well-being of individuals, corporations, and governments. Therefore, understanding the neurocognitive processes that lead individuals to violate information security policy represents a fundamental pursuit for NeuroIS researchers. In the current study, we examined the effects of whether an individual or a close associate benefited from a violation of information security, and the temporal delay before the benefit was received on event-related brain potentials (ERPs) related to ethical decision making. The electrophysiological data revealed modulations of the ERPs that were generally sensitive to ethical decision making, or that were specifically sensitive to the recepiant or timing of the reward. The components that were sensitive to the two independent variables were observed over the anterior frontal region of the scalp, consistent with the neuroimaging literature demonstrating that several prefrontal structures participate in self-referent processing and intertemporal choice.

**Keywords:** Information security, Ethical decision making, Event-related brain potentials

## 1. Introduction

As society has become increasingly dependent upon digital information, the impact of cybercrime has increased exponentially. Cybercrime may reduce consumer confidence [1], create tenuous international relations [2], and is estimated to cost the world economy over three trillion dollars annually [3]. There have been significant advances in the field of computer science leading to enhancements of hardware and software technologies designed to deter cybercrime [4]. However, these advances may not

necessarily thwart the actions of individuals within an organization [5, 6], and studies demonstrate that roughly 50% of information security violations result from insider threats [7]. Unfortunately, current deterrence programs focused on information security may not reduce the intention to commit, or the incidence rate of, cybercrime arising from inside an organization [8] . Therefore, the current study builds upon recent work from our laboratory by examining the effects of two independent variables (i.e., the recipient and timing of a benefit) on event-related brain potentials (ERPs) elicited during ethical decision making related to information security. Based upon the extant literature, these two variables are known to influence decision making in various domains [9, 10].

The Information Security Paradigm (ISP) was developed by Hu et al. [11] to measure the neural correlates of ethical decision making related to information security using ERPs. This task was adapted from a survey-based research instrument used in the information systems literature. For the ISP, individuals make decisions as if they are a fictitious IT employee named Josh. In the task, participants read a series of 1-2 sentence scenarios describing violations of information security practices that vary in their degree of severity (i.e., minor or major) or control scenarios that do not involve an ethical violation. Following the scenario, subjects are presented with a decision prompt, and decide whether or not Josh should take the action described in the scenario. Comparing ethical violation scenarios to control scenarios allows one to isolate neural activity that is generally related to ethical decision making; while comparing scenarios related to minor and major violation of information security allows one to isolate neural activity that is sensitive to the severity of the violation.

Hu et al. [11] found that the behavioral and ERP data for the ISP are sensitive to both the presence and severity of ethical violations. The choice data revealed that subjects were less likely to endorse scenarios involving an ethical violation than control scenarios; while the response time data indicated that individuals considered minor violations longer than major violations. The ERP data differentiated control, minor and major ethical violation scenarios over the lateral and medial frontal regions and the right parietal region beginning around 200 ms after the onset of the prompt. In

comparison to the control scenarios, ethical violation scenarios elicited an early posterior effect on the N2, that may reflect a limit in the attentional resources available for encoding the prompt. Neural activity was also sensitive to the severity of the ethical violation. Major violations elicited greater activity over the left parietal region between 400-600 ms, revealing fairly early processing that was sensitive to the severity of the ethical violations [12]. In addition to the early activity occurring over the parietal region, there was sustained frontal-central-temporal activity that distinguished ethical violation scenarios from control scenarios that persisted for 1.5 to 2 seconds after onset of the prompt [11, 12].

In the current study, we utilized an adapted version of the ISP [11] and had two primary goals. First, we sought to provide a conceptual replication of the ERP findings related to our original materials. This goal allowed us to examine the generalizability of the behavioral and ERP data measured in the paradigm with a new set of scenarios and action prompts.

H1: There will be differences in ERP amplitude between the ethical scenarios and control scenarios that emerge beginning at around 200 ms over the occipital-temporal region and then continue over the parietal and frontal regions between 200-2000 ms.

Second, we sought to examine the effect of two independent variables (i.e., the benefactor of a reward and the timing of a reward) on the behavioral and ERP data related to decision making in the context information security. Previous research has demonstrated that the perceived benefit of a violation is a significant predictor of the intention to violate IS security policy [13], and here we sought to identify the neural basis of this effect. Additionally, the literature on intertemporal choice demonstrates that individuals are sensitive to the timing of rewards, often discounting a larger distant gain for a immediate smaller gain [9]. The functional neuroimaging literature examining the neural basis of self-referent processing and intertemporal choice has consistently revealed activation of the medial prefrontal cortex related these two constructs [9, 10]. In the ISP for the current study, the benefactor of the reward associated with the

violation was either Josh or a friend/relative; and the benefit was received after either 0-3 months or 12-24 months.

H2: Individuals will be more likely to say yes to Josh benefit scenarios than Other benefit scenarios, and to short delay scenarios that to long delay scenarios.

H3: The ERPs will reveal sustained differences in amplitude between Josh versus Other benefit scenarios, and short versus long delay scenarios, over the frontal region of the scalp.

## 2. Method

Participants. Forty individuals participated in the study, and the demographic information for one individual was lost. The participants were on average 20 years of age; and were 82% female, 79% white, 56% were Democrats and 23% were Republicans, and participants described themselves as being politically moderate ($M = 3.11$) on a 7-point scale (1 = liberal, 4 = moderate, 7 = conservative).

Materials. The Information Security Paradigm represented a 3 (benefactor: Control, Josh, Other) by 2 (timing of reward: 0-3 month delay or 12-14 month delay) factorial design with eight scenarios presented for each of the six cells of the design. Control scenarios included activities that did not involve an ethical violation; Josh scenarios involved unethical behaviors that he would benefit from; and Other scenarios involved unethical behaviors that another individual would benefit from (e.g., a friend, relative, partner) and explicitly stated the identity of the third party. The 48 scenarios were presented in a different random order for each individual. Scenarios were limited to 300 characters; and prompts were limited to 50 characters and were posed in the form of a question. The prompts did not mention the nature, benefactor, or timing of the reward. Individuals were given an unlimited amount of time to read the scenario and then pressed the spacebar to view the decision prompt. The response time and ERP data were time locked to the onset of the prompt. Individuals responded on a 4

points scale (No, Likely No, Likely Yes, Yes) using the C-V-B-N keys of the keyboard.

Procedure. After arriving at the laboratory for the study, individuals were given a brief overview of the procedure and provided signed informed consent. Individuals then completed a demographic survey and several questionnaires measuring individual differences related to self-control, moral foundations, media exposure, pathological gaming, and grit. After completing the scales, individuals were fitted with a 32 electrode actiCAP and completed the ISP, moral foundations task, and a picture rating task while EEG was recorded. Following this, individuals were debriefed and compensated with either course credit or $15.

EEG recording and analysis. The EEG was recorded from a 32 channel actiCHamp system using the Brain Vision Recorder software and a standard 32 channel actiCAP scalp montage where CP5-CP6 were replaced with active electrodes located below the eyes. During recording the electrodes were grounded to electrode Fpz and referenced to electrode Cz, for data analysis the data were re-referenced to the average reference. The EEG was digitized at 500 Hz and then bandpass filtered between .1-30 Hz using an IIR filter implemented in ERPLAB (5.1.1.0) [14] for the analyses. Ocular artifacts associated with blinks and saccades were corrected with ICA implemented in EEGLAB (13.6.5b) [15]. Trails including other artifacts were rejected before averaging using a +/- 100 □V threshold. ERPs were averaged for Control, Josh, and Other scenarios, or Short and Long scenarios from -200 to 2000 ms around onset of the prompt, and mean voltage measurements were made using the measurement tool in ERPLAB. Two to four electrodes were included in the analyses of the mean differences, with most analyses including three electrodes.

## 3. Results

Behavioral Data. The response choice and response time data were analysed in a set of 3 (scenario: Control, Josh, Other) by 2 (timing: Short or Long delay) ANOVAs (Table 1). The analysis of response choice

revealed a significant main effect of scenario, $F(2,78) = 214.45$, $p < .001$, representing a decrease in the likelihood of responding yes from Control to Josh scenarios, $t(39) = 17.68$, $p < .001$, and from Josh to Other scenarios, $t(39) = 2.58$, $p = .036$. The difference between Josh and Other scenarios provides support for Hypothesis 2. The main effect of timing was also significant, $F(1,39) = 11.52$, $p = .002$, revealing that individuals were less likely to respond yes for Short delay scenarios than Long delay scenarios; and a significant interaction, $F(2,78) = 9.63$, $p < .001$. This finding does not support Hypothesis 2. This interaction resulted from the effect of timing being significant for Josh benefit scenarios, $t(39) = 4.93$, $p < .001$, but not for Other, $t(39) = .47$, $p = .64$, or Control, $t(39) = 1.40$, $p = .17$, scenarios. The results of this analysis reveal that individuals were more likely to endorse an unethical behavior that results in a longer term personal gain.

The analysis of response time revealed a nonsignificant main effect of scenario, $F < 1.00$, and a significant main effect of timing, $F(2,39) = 4.90$, $p = .033$, revealing shorter response times for Short delay scenarios than for Long delay scenarios. The scenario by timing interaction was significant, $F(2,78) = 4.87$, $p = .01$, and resulted from shorter response times for Short than Long delay scenarios when Josh benefitted, $t(39) = 3.19$, $p = .003$, but not for Other, $t(39) = .99$, $p =.33$, or Control, $t(39)$, 1.37, $p = .18$, scenarios. The results of this analysis reveal that individuals may have thought longer about decisions related to unethical behaviors they were more likely to accept (i.e., the Josh Long delay scenarios).

Table 1. Descriptive data for choice and response time (in milliseconds) for the ISP.

| | Cont. Short | Cont. Long | Josh Short | Josh Long | Oth. Short | Oth. Long |
|---|---|---|---|---|---|---|
| Choice M | 2.88 | 2.96 | 1.60 | 1.91 | 1.64 | 1.62 |
| SD | .36 | .31 | .48 | .60 | .38 | .50 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RT | M | 2052 | 2194 | 1926 | 2290 | 2085 | 1990 |
| | SD | 778 | 613 | 706 | 1024 | 776 | 745 |

ERP Data. For the ERP data we examined three comparisons: 1) Differences between Control scenarios and Josh and Other scenarios -- collapsed across short and long delay scenarios -- were considered to identify neural activity generally related to ethical decision making. 2) Differences between Josh and Other scenarios were considered to identify neural activity related to self-referent processing. 3) Differences between Short and Long delay scenarios -- collapsed across Josh and Other scenarios -- were considered to identify the effect of temporal delay.

Table 2. Mean voltage in microvolts and omnibus F- and p-values for the comparisons of Control, Josh, and Other scenarios.

| | Occipital 200-300 | Central 200-500 | Parietal 350-600 | RT Central 350-1200 | RT Frontal 300-1000 | LT Parietal 1000-2000 |
|---|---|---|---|---|---|---|
| Control | 1.53 | -1.73 | -.01 | .77 | .83 | -.60 |
| Josh | 2.51 | -2.49 | -.91 | .04 | 1.82 | .55 |
| Other | 2.84 | -2.67 | -.60 | .07 | 2.19 | .63 |
| F, p | 6.56, .002 | 7.75, <.001 | 5.42, .006 | 8.01, <.001 | 6.04, .004 | 9.63, <.001 |

Note: Post-hoc comparisons revealed the that Control scenarios differed from Josh and Other scenarios, that did not differ from one another.

The comparison of Control scenarios versus Josh and Other scenarios provide support for Hypothesis 1, revealing differences in the ERPs between conditions beginning around 200 ms over the occipital region, that were then broadly distributed over the scalp including the parietal, central,

and frontal regions until 2000 ms after onset of the prompt (Figure 1a, Table 2). The comparison of Josh and Other trials revealed sustained ERP activity over the anterior frontal region (electrodes Fp1-Fp2, F3-F4, Figure 1b) between 300-1500 ms after onset of the prompt, $F(1,39) = 5.06$, $p = .03$, reflecting greater negativity for Other scenarios ($M = -1.12$ V) that for Josh scenarios ($M = -.47$ V). The comparison of Short and Long delay scenarios revealed sustained ERP activity over the anterior frontal region (electrodes Fp1-Fp2, Figure 1c) between 500-1500 ms after onset of the prompt that was marginally significant, $F(1,39) = 4.07$, $p = .051$, and reflected greater negativity for Long delay scenarios ($M = -.83$ V) than for Short delay scenarios ($M = .20$ V). Both of these analyses provide support for Hypothesis 3.

## 4. Discussion

The first goal of the study was to examine the generalizability of the ISP using a modified set of scenarios and prompts. Supporting Hypothesis 1, the comparison of the ERPs elicited for Control scenarios relative to Josh and Other scenarios revealed differences in ERP amplitude between control scenarios and those that included an ethical violation that were similar in time course and topography to the findings of our previous research [12]. These findings indicating that the ISP provides an assay of a core neural network involving structures within the occipital-temporal, parietal, and lateral and medial frontal cortex that contribute to ethical decision making as related to information security. Additionally, there appears to be considerable overlap between the neural correlates of ethical decision making related to information security and traditional moral reasoning tasks [16, 17]. The greater negativity over the medial frontal region for ethical scenarios relative to control scenarios may reflect conflict processing within the ACC or medial frontal cortex that arises when considering an unethical action. Extensive work using ERPs has associated the ACC with conflict processing [18], and within the ISP there is greater medial frontal negativity when individuals accept rather than reject an unethical action [19]. A finding that is consistent with the moral reasoning literature wherein the ACC is more active for difficult decisions

[17]. Slow anterior and lateral frontal ERP activity is consistently observed following more transient medial frontal activity during conflict processing in cognitive control and gambling tasks [20]; in the ISP this slow wave activity may reflect deliberative processing that reflects the weighing of the benefit to be gained against to ethical violation represented in the scenarios.

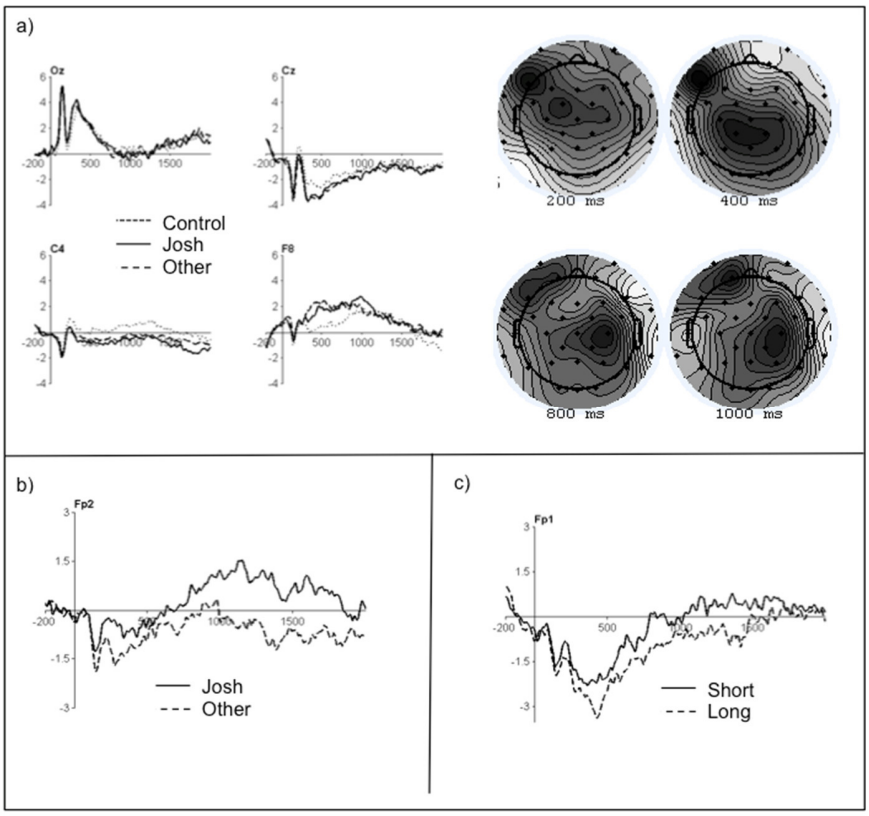



Fig. 1: a) ERPs and scalp topography maps (Josh – Control) demonstrating the timing and distributions of ERP activity that differed between ethical scenarios and control scenarios. b) ERPs demonstrating the slow frontal activity that differ for Josh vs. Other scenarios. c) ERPs demonstrating the slow frontal activity that differed for Short and Long delay scenarios.

The second goal of the study was to explore the effects of two independent variables on the neural correlates of ethical decision making in the ISP. The behavioral data provide partial support for Hypothesis 2, revealing an interaction between the independent variables that reflected a greater acceptance, and longer consideration, of long delays trials when Josh benefited relative to the other three types of scenarios involving ethical violations. These data are consistent with previous evidence demonstrating that perceived benefit is a predictor of the intention to violation information security policies [13]. Together with existing evidence, our data indicate that the benefit of a violation of information security may relate to both outcomes that might be mutually positive for the decision maker and organization (e.g., time savings) [13] or be limited to the decision maker (e.g., Josh vs. Other scenarios).

Supporting Hypothesis 3, the benefactor of the reward and the timing of the reward were associated with differences in ERP activity over the anterior frontal region between 300 ms or 500 ms and 1500 ms after onset of the prompt. The topography of the effect of these variables on the ERPs was somewhat different from those of the ERPs that distinguished Josh and Other benefit scenarios from the Control scenarios. These findings converge with the neuroimaging literature revealing that self-referent processing and intertemporal choice are consistently related to activity within the anterior frontal cortex [9, 10], and are consistent with the idea that ethical or moral reasoning arises from the recruitment of more general neurocognitive processes rather than proprietary neural circuits that are dedicated to ethical decision making [17].

There are some limitations of the study that should be acknowledged. The sample was predominantly white female undergraduates. There is continued development of the prefrontal cortex into early adulthood that may influence ethical reasoning and decision making, so it may be worthwhile in future studies to examine neural activity in the ISP in a sample in their late 20's or 30's once they entered the workforce. There is also some evidence demonstrating cultural differences in the adoption of information security practices [21], indicating that it could be useful to explore cultural variation in the ISP. Finally, we are in the process of

balancing the gender distribution within the sample to examine the possible influence of variables that may differ between males and females and that are related to information security (e.g., video game experience) [22].

In summary, the current findings demonstrate that the ISP provides a sound methodological foundation for probing the activity of a neural system that underpins ethical decision making related to information security, in addition to neural systems associated with other constructs (e.g., perceived benefit and temporal delay) that may influence decision making in this domain. Additionally, other research from our laboratory demonstrates that ERPs elicited during the ISP are sensitive to individual differences in self-control and moral beliefs [11, 12]; variables that predict the occurrence of hacking behavior [22, 23] or the intention to violation information security policy [13]. Finally, we are encouraged by the overlap in the neural systems underpinning decision making related to information security and moral reasoning more generally, and believe that this convergence has the potential to facilitate synergistic collaborations between scholars with interests in information systems, cognitive and decision neuroscience, and moral reasoning.

## References


1. Yayla, A.A., Hu, Q.: The Impact of Information Security Events on the Stock Value of Firms: The Effect of Contingency Factors. J. Inf. Technol. 26, 60-77 (2011)
2. Groll, E.: Russian Interference Went Far Beyond DNC Hack, Senate Panel Hears. Foreignpolicy.com. (2017, March 30).
3. Cybersecurity Ventures Cybereconomy Infographic (2017) http://cybersecurityventures.com/cybereconomy-infographic/
4. Ayuso, P.N., Gasca, R.M.,Lefevre, L.: A Cluster-Based Fault-Tolerant Architecture for Stateful Firewalls. Comp. Secur. 31, 524-539 (2013)

5. Posey, C., Bennett, R.J., Roberts, T. L.: Understanding the Mindset of the Abusive Insider: An Examination of Insiders' Causal Reasoning Following Internal Security Changes. Comp. Secur. 30, 486-497 (2011)
6. Warkentin, M., Wilson, R.: Behavioral and Policy Issues in Information Systems Security: The Insider Threat. Eur. J. Inform. Syst. 18, 101-105 (2009)
7. Richardson, R.: CSI Computer Crime and Security Survey. http://www.GoSCI.com (2011)
8. Hu, Q., Xu, Z., Dinev, T., Ling, H.: Does deterrence work in reducing information security policy abuse by employees? Commun. ACM. 54, 55-60 (2011)
9. Kable, J.W.: Valuation, Intertemporal Choice, and Self-control In In: Glimcher, P.W., Fehr, E. (eds.). Neuroecenomics, 2nd edition. pp. 173-192. Academic Press, London (2014)
10. Mitchell, J.P., Schirmer, J., Ames, D.L., Gilbert, D.T.: Medial Prefrontal Cortex Predicts Intertemporal Choice. J. Cognitive Neurosci. 23, 1-10 (2010)
11. Hu, Q., West, R., Smarandescu, L.: The Role of Self-control in Information Security Violations: Insights From a Cognitive Neuroscience Perspective. J. Manage. Inform. Syst. 31, 6-48 (2015)
12. Budde, E., West, R.: Neural Correlates of Ethical Decision Making Related to Information Security. Poster presented at the Midwestern Psychological Association, Chicago, IL (2017)
13. Vance, A., Siponen, M.: Is Security Policy Violations: A Rational Choice Perspective. J. Organ. End User Comput. 24, 21-41 (2012)
14. Lopez-Calderon, J., Luck, S.J.: ERPLAB: An Open-source Toolbox for the Analysis of Event-related Potentials. Front. Hum. Neurosci. 8, 213 (2014)
15. Delorme, A., Makeig, S.: EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics. J. Neurosci. Meth. 143, 9-21 (2004)
16. de Oliveira-Souza, R., Zohn, R., Moll, J.: Neural correlates of human morality: An overview. In: Decety, J., Wheatley, T. (eds.) The Moral Brain: A Multidisciplinary Perspective, pp. 183-196. The MIT Press, Cambridge, MA (2015)

17. Greene, J.D.: The Cognitive Neuroscience of Moral Judgment and Decision Making. In: Decety, J., Wheatley, T. (eds.) The Moral Brain: A Multidisciplinary Perspective, pp. 197-220. The MIT Press, Cambridge, MA (2015)
18. Cavanagh, J.F., Frank, M.J.: Frontal theta as a mechanism for cognitive control. Trends Cogn. Sci. 18, 414-421 (2014)
19. West, R., Budde, E., Malley, K.: Neural correlates of thinking about and making unethical decisions. Talk presented at the meeting of the Psychonomic Society, Vancouver, BC. (2017)
20. West, R., Bailey, K., Tiernan, B.N., Boonsuk, W., Gilbert, S.: The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. Neuropsychologia, 50, 3450-3460 (2012)
21. Dinev, T., Goo, J., Hu, Q., Nam, K.: User Behavior Towards Protective Information Technologies: The Role of National Cultural Differences. Inform. Syst. J. 19, 391-412 (2009)
22. Hu, Q., Zhang, C., Xu, Z.: Moral Beliefs, Self-control, and Sports: Effective Antidotes to the Youth Computer Hacking Epidemic. Paper presented at 45th Hawaii International Conference on Systems Science (2012)
23. Xu, Z., Hu, Q., Zhang, C.: Why computer talents become computer hackers. Comm. ACM. 56, 64-74 (2013)

# NeuroIS for Decision Support: The Case of Filmmakers and Audience Test Screenings

Sandra Pelzer[1], Marc T. P. Adam[2], and Simon Weaving[2]

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany
{sandra.pelzer}@web.de
[2] The University of Newcastle, Newcastle, Australia
{marc.adam,simon.weaving}@newcastle.edu.au

**Abstract.** The application of neuroscience theories, methods, and tools holds great potential for the development of novel decision support systems. In this paper, we develop a theoretical framework for how NeuroIS may support the test screening process of filmmakers where decisions are made about what narrative material is shown to the audience, what sequence it is to be ordered, and what emotional value it must carry. While current methods for audience test screenings commonly rely on standardized questionnaires and focus groups, decision support systems may employ neuroscience tools as built-in functions to provide the filmmaker with novel insights into how their movie is ultimately perceived by the audience. Thereby, a key focus lies on the coherence between the emotional experience intended by the filmmaker and the emotional experience exhibited by the audience. Further, NeuroIS allows an evaluation of how the emotional experience to specific cinematic moments affects overall movie satisfaction.

**Keywords:** Audience Testing, Decision Support Systems, Filmmaker, NeuroIS.

## 1 Introduction

Over the past ten years, the application of neuroscience theories, methods, and tools has provided valuable theoretical and methodological insights for information systems research, particularly in terms of informing the design of IT artifacts and using neuroscience tools for their evaluation. However, only few studies have explored how biosignals can be used as built-in functions of IT artifacts such as decision support systems [1, 2]. In this paper, we explore this promising path of design science research by developing a theoretical framework for how NeuroIS tools may support filmmakers in the process of finishing a film for distribution, when a series of decisions are made in post-production that finally determine what narrative material is shown to the audience, what sequence it is to be ordered, and what emotional value it must carry. The framework enables filmmakers (1) to evaluate the level of coherence between the filmmaker's intentions for the emotional experience at specific moments of visual storytelling and the audience's exhibited emotional experience and (2) to

identify how the emotional experiences in response to specific cinematic moments affect overall movie satisfaction.

The global movie box-office for 2017 reached US$40 billion, with movies playing in 125,000 screens in more than 25,000 cinemas across the world [3]. Yet, despite estimates of total global movie production exceeding more than 3,000 films a year [4], the industry is characterised by "high stakes, highly uncertain investments" [5] and high failure rates of individual movies. In addition, industry practitioners rely upon "tradition, conventional wisdom and simple rules of thumb" [6] rather than more scientific approaches to creative and managerial decision-making. In particular, knowledge of the emotional experience of the audience and its link to the success of a movie is scarce, with audience testing typically limited to standardized questionnaires and focus groups to follow up and find more detailed qualitative causes. Feedback is focused primarily upon ascertaining an overall rating (e.g., "Would you recommend this movie to your friends?") and consideration of pre-selected aspects of the film thought to be potentially problematic (e.g., concerns over a main character's likeability). Results from audience testing are often aimed at discovering elements for marketing campaigns, but also provide information so that the filmmaker can make adjustments in the final stages of the post-production process [7].

While the existing approaches of questionnaires and focus groups provide important insights into an audience's overall perception of a movie, they allow for little exploration of the perception of individual cinematic moments at an emotional level. As complex forms of storytelling, movies are developed through screenwriting, brought to life by direction, and finally constructed with editing. Throughout this interconnected process, the intended emotional response of the audience is the primary concern, particularly for the roles of screenwriter, director, and editor. Whatever the genre, movies are designed to give pleasure by provoking emotion: to be successful horror films must scare, thrillers must thrill, and "weepies" must make us cry. For screenwriters, "what we are really after, what we are really concerned about is the emotion […] What is the emotion underpinning the scene, this story?" [7, p. 25]. Directing involves "a passion for the human condition and characters and their emotional state of mind from moment to moment" [8, p. 3], whilst the art of editing places the highest value on being "true to the emotion of the moment" [9, p. 18].

In this paper, we elaborate on how utilizing biosignals may support the decision making of filmmakers in the post-production process. Thereby, we build on the recent advances in NeuroIS research in employing neurophysiological measurements such as electroencephalography (EEG), heart rate variability, skin conductance, and startle reflex modulation as measures for human affective processing [10, 11, 12], and the integration of these measures as built-in functions of IT artifacts [1, 2, 13].

## 2 Theoretical Framework

Traditional approaches for gathering feedback from audiences prior to the release of a movie generally focus on the audience's perception of the movie as a whole. However, the making of a movie involves a multitude of decisions around how the

narrative is to be delivered by means of a complex, and yet sequential, set of audio-visual sensory stimuli. Hence, an audience's overall perception of a movie is a function of how they experience this sequence of audio-visual sensory stimuli, and the way this experience leads to a re-construction of the story world in the mind of the audience. The development of our theoretical framework starts from the rationale that neurophysiological measurements may provide important insights for the filmmaker into how their selection of audio-visual stimuli is ultimately experienced by the audience, and how the experience of individual segments is reflected in their overall movie satisfaction.

Importantly, our framework particularly focuses on those segments of a movie that the filmmaker believes to play a critical role in the perception of a movie, referred to in the following as *cinematic moments*. While a cinematic moment may last anywhere between only a few seconds to several minutes, it draws its significance from the meaning that the filmmaker intends against the backdrop of the plot and the story's characters. In most instances, the meaning inherent to a cinematic moment is carried at least partially by an emotional experience (e.g., anger, relief) intended by the filmmaker. Building on this rationale, our proposed theoretical framework sees the concept of cinematic moments and the emotional experience created through them as antecedents of overall movie satisfaction (see Fig. 1).

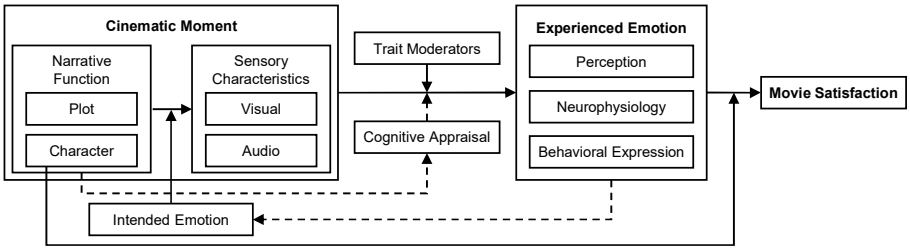



**Fig. 1.** Theoretical Framework

The framework establishes the notion that a filmmaker may induce emotions through two groups of interrelated components. First, the *narrative function* refers to the meaning that the cinematic moment carries for plot and/or the characters of the film. For instance, the moment when Mrs. Bates (Norman's mother) is revealed to be a skeleton at the end of the movie *Psycho* (1960), both solves the plot's mystery of who the killer is, and makes us re-evaluate the character of Norman Bates (all accomplished while provoking shock and terror in the audience). Second, *sensory characteristics* refer to the specific auditory and visual elements that are chosen by the filmmaker to fulfil the narrative function of the cinematic moment. Filmmakers make choices about what we see and hear, and these choices are not only made to maximise narrative comprehension but have an intended emotional response in mind. For instance, in the skeleton reveal scene, director Alfred Hitchcock chooses to cut to a close up just as the skeletal face of Mrs. Bates is revealed, and punctuates the moment with a scream and the dramatic violin-dominated theme music.

Further, we conceptualize that the emotional experience that the filmmaker intends to invoke in the audience (intended emotion) ultimately leads to an actual emotional experience in the audience (experienced emotion) as conveyed by the sensory characteristics of the cinematic moment. The experience of emotions depends on the cognitive appraisal of the narrative function of the cinematic moment (e.g., confusion might be a desirable emotion for the climax of a psychological thriller, but not for establishing the story world of a drama) as well as individual characteristics of the audience (e.g., movie preferences, expectations of genre). Thereby, each experienced emotion may comprise perception, neurophysiological activation patterns, and behavioural expressions. The set of intended emotions is revealed in the sensory elements a filmmaker chooses when delivering the narrative function of a cinematic moment. We therefore conceptualize that the way narrative functions are delivered by audio-visual elements is moderated by a filmmaker's intended emotions. Further, even though a cinematic moment may trigger a range of emotions in the audience, this emotional experience may not necessarily be in tune with what the filmmaker intended.

Based on the conceptualization of cinematic moments, and their interplay with the audience's emotional experience and overall movie satisfaction, the framework enables us to examine a range of relationships between the investigated constructs (e.g., intended emotion, experience emotion, movie satisfaction) that may provide important insights for the movie production process. For instance, it enables the filmmaker to see (1) the coherence of emotions experienced by an audience and the emotions intended by the filmmaker for each cinematic moment, (2) the degree of divergence in emotional experience across different members of the audience (e.g., if some audience members are bored by a long action sequence, whilst others find it thrilling), and (3) the way in which the emotional experience to individual cinematic moments contribute towards overall movie satisfaction.

We posit that in order to effectively operationalize the measurement of the pathways expressed in the proposed theoretical framework, neurophysiological measurements provide a promising avenue for audience test screenings. Combined with self-report data on the nature of the emotional state and the audience's overall satisfaction with the movie, neurophysiological measurements allow for the collection of information on how an audience experiences individual cinematic moments without interrupting the overall movie experience. Previous research has involved the analysis of movie preferences using neurophysiological signals (e.g., EEG [14], heart rate [15]), but these approaches were not intended to provide decision support for filmmakers, whose active involvement in the selection of the cinematic moments, and the definition of the intended emotional experience for each, form the criteria around which measurement occurs. Operationalizing the measurements this way can provide detailed decision-support for filmmakers prior to distribution, with the aim of reducing risks of failure in a highly competitive global market.

In particular, over the past decade NeuroIS scholars have explored a range of neurophysiological measures to investigate human affective processing in human-computer interaction. Thereby, a common approach is to follow a dimensional perspective of emotion as expressed in Russel's circumplex model of affect [16], which considers the dimensions of hedonic *valence* (from unpleasant to pleasant) and *arous-*

*al* (from unaroused to aroused) as key aspects of a person's emotional state. As for the valence dimension, scholars have employed measures such as EEG (e.g., to predict e-loyalty in websites [10]) and startle reflex modulation (e.g., to predict attitudes towards brands and virtual reality [12]). Complementarily, measures such as skin conductance response and certain aspects in heart rate variability (e.g., in electronic auctions [11], 13]) have been employed to assess the arousal dimensions of users' emotional experience. Further, eye tracking has been used to synchronize neurophysiological activity with the time at which a user perceives a certain stimulus on the screen (e.g., users' responses to email pop-up notifications [17]). Finally, several studies have explored how such biosignals may be used as real-time system input for information systems (e.g., for stress management [2] and emotion regulation [13]). Applied to the case of decision supports systems for filmmakers, such measures could be used in audience test screenings to provide insight into the emotions an audience experiences during cinematic moments identified by the filmmaker without interrupting the film. The system may then contrast these data with the emotional experience intended by the filmmaker, and compare them with data collected using follow up questionnaires on the audience's perceptions of specific cinematic moments after the end of the film (e.g., identification of emotional states).

## 3 Discussion and Future Work

The theoretical framework presented in this paper may provide a first step towards employing NeuroIS tools for decision support in audience test screenings. At this stage, filmmakers have little information as to how the emotional experiences they intend to create through audio-visual stimuli lead to overall satisfaction with the movie. The theoretical framework allows us to identify key constructs in the perception of a movie that drive movie satisfaction, which in turn enable us to devise operationalizations with neurophysiological measurements for the design of decision support systems. Conceptually, such a system will enable a feedback loop between a filmmaker's intentions in the making of a movie, and the audience's actual experience – information that may turn out critical for decision making in post-production.

Building on the proposed framework, a proof-of-concept has been implemented using the software platform *Brownie* [18, 19]. Its main purpose was to conduct an initial test of the framework using a nearly completed short film where a number of cinematic moments and their intended emotions were identified by the filmmaker, and compared to the emotional experience of a small group of viewers. Because an audience's satisfaction with a movie is also subject to their expectation towards that movie [20], the proof-of-concept deliberately avoided setting specific expectations about the film experience. The results provide support for our rationale that the emotional experience and satisfaction for individual cinematic moments contributes to the audience's overall movie satisfaction. Further, a higher degree of convergence between a filmmaker's intention and the audience's actual emotional experience is also associated with a higher degree of movie satisfaction, which supports our hypothesis that if an audience member does not feel the intended emotions with the intended intensity, the

satisfaction with a moment or the movie as a whole drops. Finally, a higher degree of divergence in emotional experience across the audience is related to a lower degree of overall movie satisfaction.

We suggest further research in a number of areas. The first is in identifying a set of particular emotional states that are most relevant for the movie watching experience and the link between these and the overall satisfaction of a movie experience. This set of emotional states is essential to support the filmmaker in selecting cinematic moments and defining the intended emotional experience. Secondly, as "overall satisfaction" is not a discrete emotion but presumably formed as part of the emotional journey experienced, further research would be useful in establishing the links between the extent to which specific emotions are experienced as cinematic moments unfold, and the overall sensation of being "satisfied" with the whole movie. Thirdly, we suggest research into how neurophysiological data can be used in decision support systems to forecast satisfaction with a more diverse range of genres (e.g., comedy, drama) and with creative content beyond traditional feature films (e.g., documentaries, music videos). Such research would be a highly positive step for filmmakers, who are in search of meaningful tools to replace the rules of thumb and long-established processes that dominate decision-making in their highly competitive and risky business environment.

## References


1. vom Brocke, J., Riedl, R., Léger, P.-M.: Application strategies for neuroscience in information systems design science research. J. Comput. Inf. Syst. 53, 1–13 (2013).
2. Adam, M.T.P., Gimpel, H., Maedche, A., Riedl, R.: Design blueprint for stress-sensitive adaptive enterprise systems. Bus. Inf. Syst. Eng. 59, 277–291 (2017).
3. Comscore: Comscore Reports Worldwide Box-Office Hits a New All-Time Record, https://ir.comscore.com/news-releases/news-release-details/comscore-reports-worldwide-box-office-hits-new-all-time-record.
4. The numbers: Movies released in 2017, https://www.the-numbers.com/movies/year/2017 (2017).
5. Chisholm, D., Fernandez-Blanco, V., Ravid, S., Walls, W.: Economics of motion pictures: the state of the art. J. Cult. Econ. 39, 1–13 (2015).
6. Eliashberg, J., Elberse, A., Leenders, M.: The motion picture industry: Critical issues in practice, current research & new research directions. Mark. Sci. 25, 638–661 (2005).
7. Bauer, I.: Screenwriting Fundamentals: The Art and Craft of Visual Writing. Routledge, New York / London (2017).
8. Schreibman, M.: The film director prepares: a practical guide to directing for film and TV. Lone Eagle Publishing, New York (2006).
9. Murch, W.: In The Blink of an Eye. A Perspective on Film Editing. Silman-James Press, Hollywood (1992).
10. Gregor, S., Lin, A.C.H., Gedeon, T., Riaz, A.: Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research. J. Manag. Inf. Syst. 30, 13–48 (2014).
11. Teubner, T., Adam, M.T.P., Riordan, R.: The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions. J. Assoc. Inf.

Syst. 16, 838–879 (2015).
12. Koller, M., Walla, P.: Measuring affective information processing in information systems and consumer research: Introducing startle reflex modulation. In: ICIS 2011 Proceedings, pp. 1129-1140, Shanghai, China (2011).
13. Astor, P.J., Adam, M.T.P., Jerčić, P., Schaaff, K., Weinhardt, C.: Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. J. Manag. Inf. Syst. 30, 247–278 (2013).
14. Dmochowski, J.P., Bezdek, M.A., Abelson, B.P., Johnson, J.S., Schumacher, E.H., Parra, L.C.: Audience preferences are predicted by temporal reliability of neural processing. Nat. Commun. 5, 1–9 (2014).
15. Christoforou, C., Christou-Champi, S., Constantinidou, F., Theodorou, M.: From the eyes and the heart: A novel eye-gaze metric that predicts video preferences of a large audience. Front. Psychol. 6, 1–11 (2015).
16. Russel, J.A.: A circumplex model of affect. J. Pers. Soc. Psyc. 39(6), 1161–1178 (1980).
17. Léger, P.-M., Sénecal, S., Courtemanche, F., Ortiz de Guinea, A., Titah, R., Fredette, M., Labonte-LeMoyne, É: Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research. J. Assoc. Inf. Syst. 15, 651–678 (2014).
18. Hariharan, A., Adam, M.T.P., Dorner, V., Lux, E., Müller, M.B., Pfeiffer, J., Weinhardt, C.: Brownie: A platform for conducting NeuroIS experiments. J. Assoc. Inf. Syst. 18, 264–296 (2016).
19. Jung, D., Adam, M.T.P., Dorner, V., Hariharan, A.: A Practical Guide for Human Lab Experiments in Information Systems Research: A Tutorial with Brownie. J. Syst. Inf. Technol. 19, 228–256 (2017).
20. Aurier, P., Guintcheva, G.: Using affect-expectations theory to explain the direction of the impacts of experiential emotions on satisfaction. Psyc. Mark. 31, 900–913 (2014).

# Measuring the Impact of Mind Wandering in Real Time Using an Auditory Evoked Potential

Colin Conrad and Aaron Newman[1]

[1]Dalhousie University, Halifax, Canada
{colin.conarad, aaron.newman}@dal.ca

**Abstract.** In this research-in-progress paper, we propose an experiment to investigate the neurophysiological correlates of mind wandering using electroencephalography (EEG). Auditory oddball event related potentials have been observed to be sensitive to the mind wandering state and can be used as a real-time passive measure. This has advantages over standard survey techniques because it is an objective, non-disruptive real time measure. We describe an experiment to observe the neurophysiological correlates of mind wandering in online learning environments using an auditory oddball. In doing so, we introduce a new experimental paradigm to the IS literature which could be used to extend other attention-related research.

**Keywords:** Auditory oddball · Mind wandering · Online learning · EEG

## 1. Introduction

Mind wandering refers to processes commonly described as "daydreaming," or "spontaneous thoughts" [1]. More precisely, mind wandering represents a phenomenon where sustained attention is brought away from a stimulus and toward self-generated experiences [2]. It is commonly thought that mind wandering occurs in the higher education environment, and though it varies from student to student, it is often perceived to have an overall negative impact on student performance [3, 4]. In the case of common online learning systems such as Massive Open Online Courses (MOOCs), it is tempting to make similar inferences, as they often likewise follow a lecture format. One key difference between the classroom and the online lecture format however, is that good classroom teachers can often observe behaviors characteristic of mind wandering and improve their teaching to increase engagement. Detecting mind wandering in an online learning environment would be useful to improving e-learning systems and identifying improved methods for objectively measuring mind wandering would be instrumental to the improvement of such systems.

In order to measure the impact of mind wandering on education, we explore using two electroencephalography (EEG) measures. The first measure is commonly referred to as the P1-N1-P2 auditory event related potential (ERP), which consists of a sequence of three peaks that consistently appear in response to the onset of auditory

stimuli, with characteristic timing and scalp distributions [5]. Studies in mind wandering have found an effect where the amplitude of the P2 elicited by auditory oddball stimuli is reduced in individuals who have attention directed away from task-relevant stimuli and toward self-generated information [6, 2]. The second measure consists of oscillatory patterns in specific frequency bands, commonly referred to as delta, theta and alpha activity, which have been found to be correlated with mind wandering [6]. In this research-in-progress paper, we describe an experiment to identify the differences in these two patterns and their correlation with self-reported mind wandering. We propose employing these methods to conduct research in real-time changes in covert attention, which are relevant to predicting performance in online learning.

## 2. Hypothesis Development

Mind wandering is a common phenomenon that plays a significant role in general thought processes, even taking up to 50% of our waking time [7]. Mind wandering is also understood to be associated self-generated thoughts and with the default mode network, which is the series of mental functions active in the absence of an explicit task. The activation of self-generated thought processes carry both costs and benefits from the perspective of cognition, depending on the context in which they are active. Self-generated thoughts have been observed to contribute to absentmindedness and unhappiness, but also have the benefit of facilitating creativity and planning [8].

Though self-generated thoughts seem to play an essential role in common human experience, the role they play in learning is inconclusive. In the context of information technology, Sullivan, Davis and Koh performed exploratory work on this subject and found that not all types of mind wandering are detrimental to learning and some forms might in fact be beneficial [9]. However, other studies affirm its overall negative impact on learning. In a study of 463 undergraduate students, Lindquist and McLean found that students who experienced frequent task-unrelated images and thoughts performed poorer in course examinations and that experiencing task-unrelated thoughts was negatively correlated with degree of course interest [4]. Mind wandering has also been found to be correlated with the activation of brain regions associated with cognitive control and executive networks, and may even compete for resources with learning stimuli [10]. Though it far from conclusive, we can hypothesize that mind wandering is generally detrimental to knowledge acquisition, at least when it comes to the sorts of knowledge acquired with executive networks, such as rote learning.

H1 – Reported mind wandering will be negatively correlated with rote learning.

### 2.1 Measuring Mind Wandering Using Neurophysiological Indicators

Though mind wandering can be effectively measured using ex post questionnaires, these methods do not offer insight into the temporal impact of mind wandering. It is desirable to develop measures that can offer insight on the changes in mind wandering

patterns over time, as temporal data can help identify which portions of an online learning system account for changes in mind wandering patterns. One method for doing this is experience sampling, a series of very short self-reports designed to capture the temporal experience of participants. Studies using these methods often employ a simple yes/no measure in order to determine the occurrence of mind wandering [11, 12]. This comes with the advantage of measuring mind wandering in real time, but with the disadvantage of disrupting the person's current cognitive processes, be they task-related or mind wandering.

Neuroimaging can be used to mitigate this problem. Oddball protocols can be used to elicit event-related potential responses from a given stimulus during a sustained task such as an e-learning session and have already been demonstrated in the IS context [13]. The P1-N1-P2 complex is a series of event related potentials triggered by an auditory or visual stimulus and can be adapted to this task. Established by Hillyard, Vogel and Luck, this complex is a mandatory response triggered by early attention control mechanisms in the occipital regions [5]. The P1-N-P2 complex has been found to be a significant indicator of the switch of general selective attention from one stimulus to another, most notably by differences in amplitude between attended and ignored stimuli. The amplitude of the P2 component was also observed to be sensitive to mind wandering by Braboszcz and Delorme [6]. Using an passive auditory oddball protocol, they demonstrated significant differences between the P2 amplitudes between participants reporting to be in a mind-wandering state versus on task.

In addition to the P2 response, correlations between oscillatory activity and mind wandering have been found at the delta, theta, alpha and beta bands [6]. Neural oscillations are caused by neural activity in the central nervous system and underline at least two modes if cerebral activity: fast-frequency waves reflective of high degrees of task-related attention (beta activity at 12-30 Hz) and a low-frequency waves reflective of low task-related attention (theta activity at 3-7 Hz). Braboszcz and Delorme also observed the impact of oscillatory activity on mind wandering ultimately found theta and beta to be significant correlates of mind wandering, while noting that delta and alpha activity was suggestive. We are led to the following hypotheses:

H2a – Mean P2 amplitude will be positively correlated with reported mind wandering.
H2b – Delta power will be positively correlated with reported mind wandering.
H2c – Theta power will be positively correlated with reported mind wandering.
H2d – Alpha power will be positively correlated with reported mind wandering.
H2e – Beta power will be negatively correlated with reported mind wandering.

## 3. Experiment Design

Participants will be asked to attend to a 51-minute English language video on Machine Learning as auditory stimuli are presented [14]. The subject matter and video were chosen because the subject matter is not commonly taught in the standard business curriculum, had some utility to the participants, and was observed triggering

variations in mind wandering during pilot studies. The video consists of a standard lecture along with a visual aid created in Microsoft PowerPoint. Participants are asked to pay attention to the video, while being presented with one of two audio stimuli every 1-1.5 seconds (mean 1.25). Participants are asked to report when they experience mind wondering by pushing a button on the computer keyboard, which is recorded on the parallel port. Following the video, participants complete a multiple-choice quiz to measure retention. Participants also complete a short multiple-choice quiz before and after the video. The differences in results are used as a measure of rote learning.

### 3.1 Participants

Twenty-four healthy students between the ages of 19 and 29 will be recruited from Dalhousie University to participate in the study. Power analysis on the oddball response suggest that this number would be for 99% confidence. Participants will be screened for neurophysiological, emotional, medical, hearing and vision conditions that could lead to abnormal EEG. Participants will also be excluded if they are majoring in computer science, have taken a course related to machine learning or are not fluent in English. Participants will be compensated CAD $25 for their time.

### 3.2 Experimental Stimuli

All stimuli will be presented in a controlled computer environment in a small, quiet testing room. Audio stimuli consist of 100 ms tones delivered every 1-3 s (randomly distributed with mean of 2 s). Task standard stimuli (80% of trials) consist of 500 Hz tones while the oddball (20% of trials, pseudo-randomly distributed) stimuli are 1000 Hz. Exactly 2448 tones are presented in the course of the experiment. The PsychoPy Python library is used to present the audio stimuli and record manual responses [15]. The onset of each audio tone is communicated to the EEG amplifier via TTL codes sent from PsychoPy via the parallel port.

### 3.3 Procedure

After completing the informed consent procedure, participants are fitted with the EEG cap (see next section) and brought to the controlled environment. Participants watch the 51 minute machine learning video, and are asked to press a button on the computer keyboard every time they become aware that their mind is wandering. Following the study, participants complete a multiple-choice quiz to test their retention of the material presented in the video.

### 3.4 EEG Data Acquisition

Participants are fitted with 32-channel scalp electrodes (ActiCap, BrainProducts GmbH, Munich, Germany) positioned at standard locations according to the International 10-10 system, and referenced during recording to the midline frontal (FCz)

location. Bipolar recordings are made between the outer canthi of the two eyes, and above and below one eye, to monitor for eye movements and blinks. Electrode impedances are kept below 15 kOhm throughout the experiment. EEG data are sampled at 512 Hz using a Refa8 amplifier (ANT, Enschende, The Netherlands), bandpass filtered between 0.01 and 170 Hz, and saved digitally using ASAlab software (ANT).

### 3.5 Artifacts Correction and Data Processing

The MNE-Python library [16] is used for data preprocessing. A 0.1–40 Hz bandpass filter is applied to the data, followed by manual identification and removal of electrodes and epochs with excessive noise. The data are then segmented into epochs spanning 200 ms prior to the onset of each auditory tone, to 1 s after. Independent Components Analysis is then used to identify and remove artifacts such as eye blinks and movements [17]. The epochs that occur in the 10 s before the reported mind wandering (excluding the 1 s window before the report) are assigned a “mind wandering” label, while epochs that occur in the 10 s after the reported mind wandering (excluding the 1 s window after the report) are assigned an “on-task” label. Fig. 1 illustrates how the data are prepared for analysis.

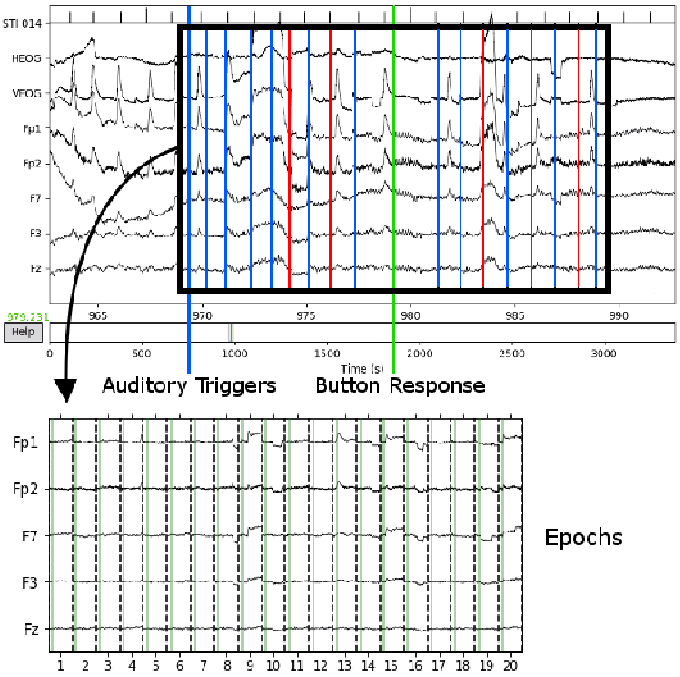



**Fig. 1.** Auditory events are triggered in PsychoPy and recorded in the parallel port. Though thousands of events are recorded, only the 1.2 s epochs from the auditory events in the 10 seconds before the button response ('mind wandering condition') and the 10 seconds following the response ('on task condition') are compared.

Planned comparisons are between standard and oddball stimuli, within and between mind wandering and on-task conditions. Pilot results (n=11; see below) suggest a high variance in mind wandering reports among participants, ranging from 1 to 60 responses. Following the recommendations of Braboszcz and Delorme [6], participants with fewer than 20 oddball responses will be excluded. Each participant is expected to yield between 20 and 140 mind wandering or on-task oddball events. In addition to temporal domain (ERP) analyses of the P1-N1-P2 components, time-frequency analysis will be investigated in the 10 s pre- and post-report. These longer epochs will assessed for power spectral density ($uV^2/Hz$) in each standard EEG frequency band.

## 4. Pilot Study and Future Work

We conducted a pilot study of this paradigm with 11 participants. Of the 11 participants recruited, 3 had to be excluded due to technical errors or lack of mind wandering measures. After data processing there were 2251 standard and 474 oddball epochs with the "on task" label, and 1887 standard and 417 oddball usable epochs with the "mind wandering" label. Fig. 2 visualizes the differences in the grand average between the standard and oddball ERP and the two conditions.

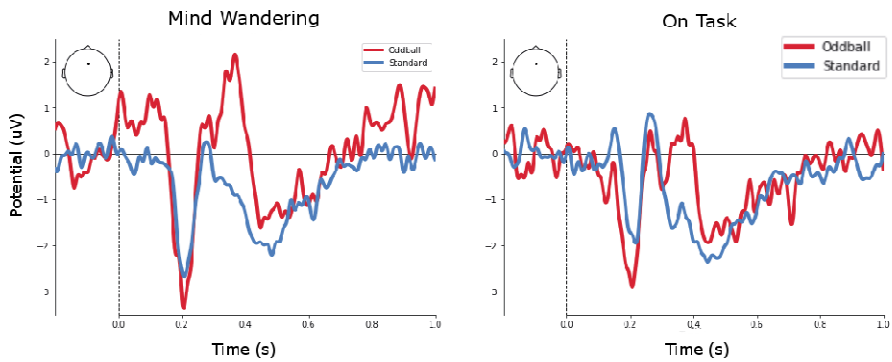



**Fig. 2.** Grand average ERP observed during mind wandering and on task conditions for channel Fz

In both mind wandering and on-task conditions, clear differences were observed between standard and oddball stimuli over midline frontal electrodes at two times: at approximately 200 ms—with a greater negativity for oddballs—and from approximately 300-400 ms—with oddball stimuli showing a more positive amplitude over midline frontal electrodes. These correspond to the typically described N1 and P3 components, respectively. Though the N1 effect appears similar to that observed by Braboszcz and Delorme [6], the enhanced positivity occurs on the P3 component, rather than on the P2 as reported by Braboszcz and Delorme. The P3 is commonly elicited by oddball stimuli in paradigms such as this, however it is more commonly associated with task-relevant stimuli—whereas here the stimuli were to be ignored.

Interestingly however, the P3 appears larger in the present data during the mind wandering thank on-task periods. We speculate that this could be caused by participants' paying greater attention to the auditory stimuli when their attention was less focused on the video (i.e., during mind wandering) the auditory stimuli drawing attention away from the video to a greater degree in the mind wandering state. As these were pilot data no statistical analyses were performed, but linear mixed effects analysis will be used once the full sample has been collected.

These preliminary results provide encouraging support for the proposal that this paradigm represents an automatic, covert, and temporally sensitive measure of mind wandering that can be applied in a range of task settings. If the auditory oddball correlates of mind-wandering are successfully established for online learning research, we can envision extending this measure to answer questions about the role of mind wandering in other technology environments. This could complement other psychophysiological measures such as eye movements or electrodermal activity, which could in turn be used to investigate the role of mind wandering outside of human-computer interaction, such as in-group dynamics or conversation [18,19]. Additionally, a robust understanding of these correlates open up the potential of attention-adaptive interfaces, which have applications to information technology generally.

**Acknowledgements.** This research is supported by the Killam and NSERC Doctoral scholarships to Colin Conrad, and an NSERC Discovery Grant to Aaron Newman. We also thank the participants of the 2017 NeuroIS training course for their feedback.

## References

1. Christoff, K., Gordon, A., and Smith, R.: The role of spontaneous thought in human cognition. In: Vartanian, O., Mandel, D. (eds.) Neuroscience of decision making, pp. 259-84. Psychology Press, New York (2011).
2. Smallwood, J.: Mind Wandering and Attention. In: Fawcett, J., Risk, E., and Kingstone, A. (eds.) The Handbook of Attention, pp. 233-255. MIT Press, Cambridge (2015).
3. Wilson, K. and Korn, J. H.: Attention during lectures: beyond ten minutes. In: Teaching of Psychology vol. 34, no. 2, pp. 85-89 (2007).
4. Lindquest, S. I. and McLean, J. P. Daydreaming and its correlates in an educational environment. In: Learning and Individual Differences vol. 21, no 2, pp. 158-167 (2011).
5. Hillyard, S. A., Vogel, E. K., and Luck, S. J.: Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. In: Philosophical Transactions of the Royal Society of London B: Biological Sciences vol. 353, no. 1373, pp. 1257-1270 (1998).
6. Braboszcz, C., and Delorme, A.: Lost in thoughts: neural markers of low alertness during mind wandering. In: Neuroimage vol. 54, no. 3, pp. 3040-3047 (2011).
7. Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., and Sayette, M. A.: Meta-awareness, perceptual decoupling and the wandering mind. In: Trends in cognitive science vol. 15(7), pp. 319-326 (2011).

8. Smallwood, J., and Andrews-Hanna, J.: Not all minds that wander are lost: the importance of a balanced perspective on the mind wandering state. In: Frontiers of Psychology vol. 4, no. 441 (2013).
9. Sullivan, Y., Davis, F., and Koh, C.: Exploring Mind Wandering in a Technological Setting. In: Proceedings of the 36th International Conference on Information Systems, Fort Worth (2015).
10. Moss, J., Schunn, C. D., Schneider, W., and McNamara, D. S.: The nature of mind wandering during reading varies with the cognitive control demands of the reading strategy. In: Brain Research vol. 1539, pp. 48-60 (2013).
11. Smallwood, J. and Schooler, J. W.: The restless mind. In: Psychological Bulletin vol. 132 no. 6, pp 946-958 (2006).
12. Schooler, J. W., Reichle, E. D., and Halpern D. V.: Zoning Out while Reading: Evidence for Dissociations between Experience and Metaconsciousness. In: Levin, D. T. (ed.) Thinking and Seeing: Visual metacognition in adults and children, pp. 204-226. MIT Press (2004).
13. Léger, P. M., Sénecal, S., Courtemanche, F., Oritz de Guinea, A., Titah, R., Fredette, M., Labonte-LeMoyne, E.: Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research. In: Journal of the Association for Information Systems vol. 15, no. 10, pp 651-678 (2014).
14. Grimson, E. Introduction to Machine Learning. YouTube (2017).
15. Peirce, J. W.: Generating stimuli for neuroscience using PsychoPy. In: Frontiers in neuroinformatics vol. 2, no. 10 (2009).
16. Gramfort, A., Luessi, M., Larson, e., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M. S.: MNE software for processing MEG and EEG data. In: Neuroimage vol. 86, pp. 446-460 (2014).
17. Delorme, A. & Makeig, S.:. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. In: Journal of Neuroscience Methods vol. 134, no. 1, pp. 9–21 (2004).
18. Douglas, K., Tremblay, A., and Newman, A.: Measuring the N400 during naturalistic conversation: An EEG hyperscanning study. Poster to be presented at the Neurobiology of Language meeting, Quebec City, August 18-22 (2018).
19. Tremblay, A., Flick, G., Blanchard, A., Cochingyan, E., Dickie, K., Macgillivray, K., Ashraf Mahmoud-Ahmed, H., Schwartz, S., Sher, C., Asp, E., and Newman, A.J.: Free, Unscripted Conversation with MEG Recordings is a Viable Experimental Paradigm for Research into Language Processing. Paper presented at The 40th Atlantic Provinces Linguistic Association (APLA-40), October 28-29, 2016, Mount Saint Vincent University, Halifax, Canada (2016).

# Exploring Eye-Tracking Data for Detection of Mind-wandering on Web Tasks

Jacek Gwizdka

School of Information, University of Texas at Austin, TX, USA
neurois2018@gwizdka.com

**Abstract**
Mind-wandering (MW) is a phenomenon that affects most of us; it affects our interactions with information systems. Yet the literature on its effects on human-computer interaction is only scant. This research aims to contribute to establishing eye-tracking measures that could be used to detect periods of MW while a user is engaged in interaction with online information. We conducted a lab study (N=30) and present an exploratory analysis of eye-tracking data with a focus on finding differences between periods of MW and not-MW. We found 12 eye tracking measures that were significantly different between periods of MW and not-MW. We also show promising classification results of the same variables. Our results indicate plausibility of using eye-tracking data to infer periods of MW.

**Keywords:** mind-wandering, mindless reading, eye-tracking, pupillometry.

## 1 Introduction

Most people have experienced mental state when their mind has wandered. This phenomenon is quite common and many people can remember when, for example, their reading did not result in any meaningful understanding of the text. In this case, a wandering mind can be a harmful thing as it makes us less efficient, prone to errors and to making incorrect decisions. If an information system were able to detect when a person's mind is wandering, it could offer an intervention. For example, if it detects that an e-commerce website user has spent a significant time MW while reviewing purchase options, the system could ask for additional verification before the purchase is made.

The goal of this project is to establish eye-tracking based measures of MW that could be used to detect periods of MW while a user is engaged in interaction with online information. We present an exploratory analysis of eye-tracking data (including pupillometry) with a focus on finding differences between periods of task-related and task-unrelated thoughts.

## 2 Related Work

While it is known that MW can have positive effects on human thought processes and, in particular, on creativity [1], MW is typically detrimental to the tasks that require focused attention. MW has been shown to negatively affect reading [2, 3], and the ability to resolve conflicts in displayed information [4], because the executive function is impaired. These are just two examples of negative influence of MW on user interaction with IS. MW has received increased attention from cognitive scientists and psychologists in the last decade [1, 3, 5], but we are only beginning to understand these processes.

Interestingly, it has been demonstrated that MW while reading is related to changes in eye-fixation patterns. For example, fixation durations were found to be longer and less affected by lexical and linguistic variables, while eye movements were found more erratic during MW periods than during reading [2]. Pupil dilation was found to change more spontaneously during MW periods [6, 7]. Results from these studies suggest that eye movements during MW are controlled by different cognitive processes than during normal reading. People may engage in internally focused cognitive tasks, which, presumably, are associated with different cognitive processes than externally focused tasks and unfocused MW. Recent work compared eye movements during goal-directed internally focused cognitive task and reading task. Eye behavior during the former was different and characterized, among others, by more and longer blinks, fewer microsaccades, more and shorter fixations, more saccades and saccades of higher amplitude [8]. These results suggest that some aspects of eye behavior may be coupled with internally generated information and related internal cognitive processes. We will come back to this in the Discussion section.

While MW is an important phenomenon with potentially significant explanatory power for human interaction with IS, research in this area outside psychology is still rather scant. Notable exceptions include, a theoretical model of MW in a technological setting proposed in a doctoral dissertation, with a specific goal to better understand costs and benefits of this phenomenon on technology users [9, 10]; a person-independent detection of MW based on eye-tracking data proposed by a group of computer scientists [11]; and the use of eye-tracking data and web cams in a large scale detection of MW in an education online setting [12].

Encouraged by previous research that showed relationship between episodes of MW and eye-tracking data, we aimed to 1) use more realistic stimuli than in psychology research (e.g., [6, 7]), 2) perform analysis without considering text characteristics (in contrast to [2] and to local features reported in [11]), and 3) examine differences in eye-tracking variables between periods of MW and nMW (not reported in [11], but reported in their second paper [13]).

Our research questions are as follows:

*RQ1. Which eye-tracking measures differ significantly between periods of MW and nMW?*

*RQ2. Can eye-tracking measures be used to classify periods of MW and nMW?*

## 3 Method

We conducted an eye-tracking lab experiment (N=30, 20 females) in Information eXperience (IX) lab in the School of Information at University of Texas at Austin. Eye tracking data was collected using remote eye-tracker Tobii TX-300. The experiment was approved by IRB. Each lab session typically lasted 30 minutes. At the completion of a session, each participant received $12.

### 3.1 Procedure and Materials.

Participants were pre-screened for their native or near-native level of English, and for normal to corrected-to-normal eyesight. Each participant filled out background questionnaire, performed a training task, and three online reading tasks shown in randomized order. After each task participants answered comprehension questions. These questions were included to provide motivation for attentive reading. The task design followed a simulated work-task approach [14], where tasks are presented with reasons for their performance. In our study, participants were informed that they needed to read three articles in order to prepare for a course "Technology and Society" they were taking. The articles were taken from the UBC-Hampton Reading Comprehension Test Suite [15]; their sources are listed in Table 1. Each text was presented on several web pages and was displayed in black Arial font on white background. The pages were designed to show about the same number of text lines on the page and thus each screen had about the same luminescence. Text lines height was uniform at 27px, which corresponds to 0.45° of visual angle and is approximately equal to the eye-tracker's accuracy reported by the manufacturer as 0.4°-0.5°.

**Table 1.** Articles for online reading tasks.

| |
|---|
| **1**–*A quick overview of digital activism. A blog post by Curiouscatherine (2011)* |
| **2**–*Taking the slack out of slacktivism. A popular press article published by Radio Free Europe, 2011.02.17* |
| **3**–*'Free the spectrum!' Activist encounters with old and new media technology. A journal article by Dunbar-Hester, C., published in New Media & Society* |

To capture incidents of MW, participants were periodically probed [16] and asked to indicate whether they were reading or MW [17]. We used a visual probe [18] – a pop-up window with two response buttons (Fig 1), which was displayed at random times controlled to be between 40 and 60 seconds and shown for 12 to 16 seconds at several pseudo-randomly selected locations on screen.

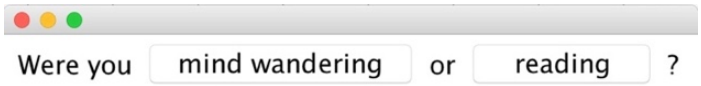



**Fig 1**. Pop-up window with the mind-wandering probe.

### 3.2 Variables

Independent variable was MW state with two levels: MW or nMW (i.e. reading). Dependent variables were obtained from eye-tracking data for two 5-second-long epochs that started respectively 5 and 10 seconds before the response to the probe. Eye-tracking data was cleaned by removing bad quality fixations (as marked by Tobii). Data from epochs with very few fixations (<3 standard deviations (std) below the mean) was discarded (<5% of epoch data) as it indicated epochs with many missing data points. We used eight types of variables: fixation count, regressions count, fixation duration, saccade duration, length, velocity, angle (four categories), and relative pupil change. For fixation duration and numerical saccade measures we calculated: mean, std, total, min and max values. For relative pupil change, which we calculated separately for each user and then denoised it with cubic interpolation of missing data (e.g., blinks), we calculated mean and std. Saccade angle was categorized to indicate (approximately) eye movement 1) forward (-10°; 10°), 2) backward (170°; 180°) in the same text line. Angles outside these ranges were categorized as 3) forward or 4) backward above or below the text line. This process yielded twenty-seven (27) variables.

## 4 Data Analysis and Results

Two participants have reported no periods of MW. Following prior work [6], we treated them as "outliers" and removed their data. Thus, we report data from 28 participants. The obtained proportion of MW responses (27.3%) matches the expectations [11, 13]. Response times to the probe were significantly longer for MW as compared with nMW periods (**Table 2**). This supports participants' correct self-classification of their internal thought processes.

**Table 2.** Responses to the probe.

| Mind-wandering | Count | % | **Response time [ms]** mean (sd) |
|---|---|---|---|
| **YES** | 115 | 27.3 | 3159 (1192) |
| **NO (reading)** | 306 | 72.7 | 2879 (933) |
| **Total** | 421 | 100 | M-WU: z=2.2, p=0.03 |

We performed inferential statistics and classification. In inferential statistics, due to not-normal distribution of variables and lack or homogeneity of variance, we used non-parametric Mann-Whitney U test (M-WU). Given the exploratory nature of our research, we conducted individual M-WU tests on each variable (**Table 3**).

We run classifications using Weka 3.8 [19]. Due to the imbalanced number of samples between MW and nMW classes, we used two data sampling methods to improve the balance, 1) synthetic sample generation SMOTE [20] and 2) random sample generation with replacement. We present best classification results obtained by applying random forest classifier with 10-fold cross-validation [21] (**Table 5**) and the best features for each classifier (**Table 4**). These features overlap with significant results in (**Table 3**).

**Table 3.** Descriptive statistics and Mann-Whitney U tests for significantly different variables.

| Epoch | Variable | # data samples (read/MW) | MW Mean(sd) | nMW Mean(sd) | M-W U |
|---|---|---|---|---|---|
| 5s | total_fixation_duration | 306/115 | 3792(761) | 3915(764) | z=2.44; p=0.015 |
| 10s | avg_fixation_duration | 288/110 | 254(87) | 238(36) | z=--2.02; p=0.043 |
| 5s | fixation_count | 306/115 | 16(3) | 16.5(3.5) | z=2.02; p=0.043 |
| 10s | - ,, - | 288/110 | 16(3) | 16.8 (3.2) | z=2.46; p=0.014 |
| 10s | tot_saccade_len | 288/110 | 2072(909) | 2220(847) | z=2.2; p=0.027 |
| 10s | max_saccade_len | 288/110 | 710(355) | 779(342) | z=1.97; p=0.049 |
| 5s | max_saccade_dur | 306/115 | 362(307) | 301(357) | z=-3.18; p=0.0015 |
| 10s | - ,, - | 288/110 | 324(311) | 298(344) | z=-2.78; p=0.0055 |
| 5s | avg_saccade_dur | 306/115 | 76(68) | 71(90) | z=-2.36; p=0.019 |
| 10s | - ,, - | 288/110 | 68(63) | 64(67) | z=-2.74; p=0.0061 |
| 5s | std_saccade_dur | 306/115 | 105(109) | 91(154) | z=-3.1; p=0.002 |
| 10s | - ,, - | 288/110 | 98(132) | 85(130) | z=-2.98; p=0.0028 |
| 10s | min_saccade_vel | 288/110 | .6(.68) | .63(2.3) | z=3.02; p=0.0025 |
| 10s | angle_cat_bck_count | 288/110 | 3.2(1.6) | 3.7(1.9) | z=2.17; p=0.03 |
| 5s | avg_pupil_change | 306/115 | -.015(.06) | -.0042(.05) | z=1.93; p=0.0535 |
| 10s | - ,, - | 288/110 | -.015(.06) | -.0019(.05) | z=2.55; p=0.011 |
| 5s | std_pupil_change | 302/109 | .034(.015) | .027(.014) | z=--4.3; p<0.0001 |

**Table 4.** Classification results – best features

| Epoch | Data sampling | **Best features** (in the order of weights from Information Gain Ranking Filter) |
|---|---|---|
| 5s | intact | std_pupil_change , std_saccade_dur |
| | SMOTE (synthetic) | angle_cat_fwd_ud_count, regression_count, std_pupil_change, std_saccade_dur, max_saccade_dur, min_saccade_vel |
| | Random with replacement | std_pupil_change, avg_saccade_dur, std_saccade_dur, max_saccade_dur, min_saccade_vel, angle_cat_fwd_count |
| 10s | intact | max_saccade_dur, std_saccade_dur, avg_saccade_dur |
| | SMOTE (synthetic) | std_saccade_dur, max_saccade_dur, avg_saccade_dur, avg_pupil_change |
| | Random with replacement | std_saccade_dur, avg_pupil_change, avg_saccade_dur, angle_cat_bck_count, max_saccade_dur, angle_cat_bck_ud_count, std_pupil_change |

## 5 Discussion

Responding to *RQ1*, we found that 12 eye-tracking measures (44% of all measures considered) significantly differed between periods of MW and nMW – seven in 5s epoch and ten in 10s epoch (in that, five measures overlapped in both epochs). These results, taken together with the confirmatory answer to *RQ2*, i.e. the reasonably promising classification results, indicate plausibility of using eye-tracking data to infer periods of MW, at least on the tasks similar to ours.

Our results generally match the previous research. For example, similarly as in [2], we found that average fixation duration (avg_fixation_duration) tended to be longer in MW periods (in 10s epochs). We also found a higher variability of changes in pupil dilation (higher avg_pupil_change and std_pupil_change) in MW periods. This is similar to results presented in [6, 7], where the authors reported more spontaneous changes

in pupil dilation during MW periods. [13] found a longer minimum saccade duration in MW periods, while we found mean and maximum saccade duration to be longer during MW, as well as a higher saccade duration variability (std_saccade_dur) during MW.

Several findings from our study are in some contrast to [8]. We found fewer fixations (in both 5s and 10s epochs) and longer average fixation duration (in 10s epochs) in MW periods, while [8] reported more and shorter fixations on their goal-directed internally focused cognitive task. This suggests that internal cognitive processes during MW periods caught in our study are different from cognitive processes during internally focused cognition described in [8]. It further suggests that the differences in cognitive processes associated with internally focused cognition, MW and reading are reflected in eye behavior and thus can be discerned from eye-tracking measures.

**Table 5.** Classification results (Random Forest with 10-fold cross validation).

| Epoch | Data sampling | Samples nMW/MW | Accuracy [%] | ROC [%] | F-measure [%] | F-measure: for MW class [%] |
|---|---|---|---|---|---|---|
| 5s | intact (no sampling) | 305/115 | 73.2 | 62.7 | 68.3 | 28.0 |
| | SMOTE (synthetic) | 306/230 | 77.4 | 84.0 | 77.3 | 72.8 |
| | Random with replacement | 210/210 | 87.6 | 96.2 | 87.6 | 87.9 |
| 10s | intact (no sampling) | 288/110 | 70.9 | 59.3 | 64.3 | 17.1 |
| | SMOTE (synthetic) | 288/220 | 76.0 | 83.2 | 75.8 | 71.1 |
| | Random with replacement | 199/199 | 89.0 | 96.3 | 88.9 | 89.1 |

Compared with [10, 12], accuracy of our classifications is higher (73%-89% in our work vs. 59%-72% and 52%-74% in their first and second work, respectively). This difference may be due to our use of different classification algorithm and a somewhat different set of features. However, our low values of F-measures for MW class (**Table 5**) indicate poor classification performance for this class when no resampling was used. This points to the need for more data samples from each user.

Contrary to prior work [2, 10], we did not find a significant difference in regression counts. Although it was marginally significant at p=.098 for 5s epoch, the difference was in the direction opposite to the expected, that is, we found fewer regressions in MW than in nWM periods. The same unexpected relationship was found in the related measure angle_cat_bck_count, which was significantly different in 10s epoch. We also have not found a higher number of line crossing saccades during MW (reported in [13]). We don't have yet a good explanation for these findings.

### 5.1 Limitations

Limitations of our work include unbalanced number of data samples from MW and nMW segments. This is expected and is due to typical frequency of MW occurrences and suggests the need to collect more data before classification algorithms are trained. We also plan to use a wider variety of epoch lengths and in the future studies, use different tasks.

## 6 Conclusion

We believe that MW is a phenomenon that will grow in importance and will be more widely studied in the context of human interaction with information systems. A broader impact of implicit detection of MW lies in its potential applicability to e-commerce and e-learning systems, where upon detection of MW an intervention could be offered to a customer or learner.

## 7 Acknowledgements

This project was supported, in part, by the Temple Teaching Fellowship 2016-2017. We thank Ms. Xueshu Chen, who was a Graduate Research Assistant, for her contributions to this project.

## References

1. Schooler, J.W., Mrazek, M.D., Franklin, M.S., Baird, B., Mooneyham, B.W., Zedelius, C., Broadway, J.M.: The Middle Way: Finding the Balance between Mindfulness and Mind-Wandering. In: Brian H. Ross (ed.) Psychology of Learning and Motivation. pp. 1–33. Academic Press (2014).
2. Reichle, E.D., Reineberg, A.E., Schooler, J.W.: Eye Movements During Mindless Reading. Psychol. Sci. 21, 1300–1310 (2010).
3. Smallwood, J.: Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention. Lang. Linguist. Compass. 5, 63–77 (2011).
4. Choi, H., Nam, C.S., Feng, J.: A Wandering Mind Cannot Resolve Conflicts in Displayed Information. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 59, 1397–1401 (2015).
5. Smallwood, J.: Distinguishing how from why the mind wanders: A process–occurrence framework for self-generated mental activity. Psychol. Bull. 139, 519–535 (2013).
6. Franklin, M.S., Broadway, J.M., Mrazek, M.D., Smallwood, J., Schooler, J.W.: Window to the wandering mind: Pupillometry of spontaneous thought while reading. Q. J. Exp. Psychol. 66, 2289–2294 (2013).
7. Smallwood, J., Brown, K.S., Tipper, C., Giesbrecht, B., Franklin, M.S., Mrazek, M.D., Carlson, J.M., Schooler, J.W.: Pupillometric Evidence for the Decoupling of Attention from Perceptual Input during Offline Thought. PLoS ONE. 6, e18298 (2011).
8. Walcher, S., Körner, C., Benedek, M.: Looking for ideas: Eye behavior during goal-directed internally focused cognition. Conscious. Cogn. 53, 165–175 (2017).
9. Sullivan, Y.: Costs and Benefits of Mind Wandering in a Technological Setting: Findings and Implications, https://digital.library.unt.edu/ark:/67531/metadc862836/.
10. Sullivan, Y., Davis, F., Koh, C.: Exploring Mind Wandering in a Technological Setting. ICIS 2015 Proc. (2015).
11. Bixler, R., D'Mello, S.: Toward Fully Automated Person-Independent Detection of Mind Wandering. In: Dimitrova, V., Kuflik, T., Chin, D., Ricci, F., Dolog, P., and Houben, G.-J. (eds.) User Modeling, Adaptation, and Personalization. pp. 37–48. Springer International Publishing (2014).

12. Zhao, Y., Lofi, C., Hauff, C.: Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach. In: Data Driven Approaches in Digital Education. pp. 330–344. Springer, Cham (2017).
13. Bixler, R., D'Mello, S.: Automatic Gaze-Based Detection of Mind Wandering with Metacognitive Awareness. In: Ricci, F., Bontcheva, K., Conlan, O., and Lawless, S. (eds.) User Modeling, Adaptation and Personalization. pp. 31–43. Springer International Publishing (2015).
14. Borlund, P.: The concept of relevance in IR. J. Am. Soc. Inf. Sci. Technol. 54, 913–925 (2003).
15. Freund, L., Kopak, R., O'Brien, H.: The effects of textual environment on reading comprehension: Implications for searching as learning. J. Inf. Sci. 42, 79–93 (2016).
16. Smallwood, J., Schooler, J.W.: The restless mind. Psychol. Bull. 132, 946–958 (2006).
17. Smallwood, J., Schooler, J.W., Turk, D.J., Cunningham, S.J., Burns, P., Macrae, C.N.: Self-reflection and the temporal focus of the wandering mind. Conscious. Cogn. 20, 1120–1126 (2011).
18. Feng, S., D'Mello, S., Graesser, A.C.: Mind wandering while reading easy and difficult texts. Psychon. Bull. Rev. 20, 586–592 (2013).
19. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2016).
20. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. ArXiv11061813 Cs. (2011).
21. Breiman, L.: Random Forests. Mach. Learn. 45, 5–32 (2001).

# Attentional Characteristics of Anomaly Detection in Conceptual Modeling

Karl-David Boutin[1], Pierre-Majorique Léger[1], Christopher J. Davis[2], Alan R. Hevner[3], and Élise Labonté-LeMoyne[1]

[1] HEC Montréal, Québec, Canada
{karl-david.boutin, pierre-majorique.leger, elise.labonte-lemoyne}@hec.ca
[2]University of South Florida, Saint Petersburg, FL, USA
davisc@mail.usf.edu
[3]University of South Florida, Tampa, FL, USA
ahevner@usf.edu

**Abstract.** We use eye tracking to better understand the attentional characteristics specific to successful error detection in conceptual models. This phase of our multi-step research project describes the visual comportments associated with successful semantic and syntactic error identification and diagnosis. We test our predictions, based on prior studies on visual attention in an error detection task, or studies comparing experts and non-experts in diverse tasks, in a controlled experiment where participants are tasked with detecting and diagnosing errors in 75 BPMN® models. The results suggest that successful error diagnostics are linked with shorter total view time and shorter fixation duration, with a significant difference between semantic and syntactic errors. By identifying the visual attention differences and tendencies associated with successful detection tasks and the investigation of semantic and syntactic errors, we highlight the non-polarity of the 'scale' of expertise and allow clear recommendations for curriculum development and training methods.

**Keywords:** eye tracking · conceptual modeling · attentional characteristics

## 1. Introduction

Business process modeling has become a central activity in IS practice [1]. Conceptual models facilitate communication about business domains and their processes [2,3,4]. Such models have become a primary medium used in design activities. This phase of our research strives to deepen understanding of visual attention during error detection tasks [5]. While researchers have explored the variations between novice and experienced modelers [6], the differences in the visual attention between successful and unsuccessful error detection tasks in conceptual modeling are yet to be deeply investigated.

For this research, we employ the Business Process Modeling Notation (BPMN®), an international standard for business processes. Its popularity in commercial settings prompted its selection for this phase of our work. Visual notations such as BPMN are composed of visual syntax - symbolic vocabulary and grammar - and visual semantics

- elements that give meaning to each symbol and symbol relationship [7,8]. However, while evaluating notations or their usage, the syntactic component is rarely discussed [8]. This presents an opportunity to contribute to the literature by comparing both the semantic and syntactic error identification process. The main objective of this study is to explore the differences in the attentional characteristics between successful and unsuccessful diagnostics in a detection task. We use eye tracking to monitor the visual attention of subjects as they search for and diagnose semantic and syntactic errors in a controlled experiment.

## 2. Prior Research and Hypotheses Development

Studies regarding the difference in the visual attention in an error detection task conclude that, on average, errors are fixated more often and longer than irrelevant information [9,10,11], and that a high number of fixations on the stimulus is correlated with an ineffective search [11,12]. Furthermore, the longer the participant spends looking for an error, the lower the chances of success [9], possibly due to too much cognitive resource being drawn away for the encoding of the stimulus. Studies that compare novices and experienced modelers point to attentional characteristics that might be associated with expertise, and thus, generally, with better performance [1],[13].

Several meta-analyses that use eye tracking to compare experts and novices in a range of domains conclude that those classified as experts spend less time looking at stimuli before fixating relevant areas or anomalies [5],[14,15,16]. More efficient scan patterns [13,14,15] or more detailed and completed schemata [17,18] are offered as explanations. Experts also tend to have fewer fixations, suggesting less cognitive effort to decipher and understand the stimuli [13],[15], and shorter fixation durations [5], which are also associated with lower cognitive processing effort [11]. Therefore, we propose three study hypotheses:

*H1 — Successful error detections in conceptual modeling will require less time spent looking at the stimulus than unsuccessful error detections.*

*H2 — Successful error detections in conceptual modeling will require, in total, fewer fixations than unsuccessful error detections, but with a higher proportion of fixations on the error.*

*H3 — Successful error detections in conceptual modeling will require, on average, shorter fixation duration than unsuccessful error detections, but with longer fixation duration on the error.*

## 3 Research Method

Our experiment was conducted on a sample of 18 participants (7 males, 11 females) with different ages and experience. All our participants were offered a $20 gift card as a compensation for their participation. The research was approved by our institution's Research Ethics Board (REB), and each participant signed a consent form.

### 3.1 Research Design and Protocol

In order to confirm our hypotheses, we tasked our participants with identifying and diagnosing errors in conceptual models written in BPMN. Each participant had to inspect five (5) distinct sets of 15 models (for a total of 75 models), where each set represented a business process scenario (e.g. airport check-in process). An example can be seen in Figure 1. Five (5) versions of each scenario were presented as BPMN 'sentences'. These were further manipulated to present three (3) versions: one with no known errors; one with a known semantic error, and one with a known syntactic error [7].

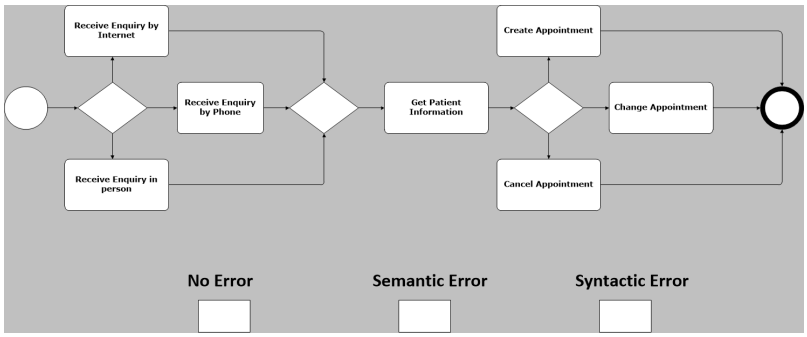



**Fig. 1.** Example of a model with boxes to indicate type of error detected

To mitigate the effect of prior knowledge of the business domains of the models [19,20], the stimuli were created using simple scenarios, well-known to all potential participants. Furthermore, we limited the range of symbols used and the scope of the 'sentences' to 10-12 elements, favoring numerous accessible models rather than more complex and domain knowledge-dependent stimuli. To train the participants and mitigate the effect of learning through the trial, the experiment started with a short presentation on BPMN [21]. The symbols used in the study, as well as the two different types of errors, were shown and explained. The training was concluded with a practice task where participants were shown three (3) models, each one with a different type of error. Just like in the real task, the participant had to pinpoint the error and to diagnose the error type, both by clicking in the corresponding area on the stimuli. The correct location of the error, as well as the right error type, was revealed after each practice model. To avoid bias, the models used in the practice task were not related to the sets of models used later in the experiment, and the conditions were the same as in the experiment [21].

The participants then started the first task with their first set of models. The fifteen models included in the set were shown in random order and without any time limit. After identifying and diagnosing the error in a model, participants had to manually advance to the next model, using the space bar on their keyboard. They then proceed to the next set of models until they completed the five (5) sets.

### 3.2 Apparatus and Measures

Eye tracking (Red 250, SensoMotoric Instruments GmbH, Teltow, Germany) was used to gather the behavioral measures throughout the experiment, at a sampling frequency of 60 Hz. The number of fixations, which is the stabilization of the eye on

an object [13], and their duration were gathered for each area of interest (AOI), as the literature tends to agree that fixation is related to the cognitive processing of visual information [13],[22]. The fixation duration threshold was set at 200 ms [11],[23]. The time before the first fixation in an AOI and the total view time of a stimulus were also collected. One (1) to three (3) AOIs were placed on the correct choice of error type and on the actual location of the error(s). For each participant, the eye tracker was calibrated to a maximum average deviation of 0.5 degrees, using a 9-points predefined calibration grid.

## 4. Preliminary Results

We briefly present several preliminary results from our study. Hypothesis 1 states that successful identification and diagnosis of errors in conceptual models will take less time than unsuccessful answers. A linear regression with mixed model and a two-tailed level of significance is performed to compare the Total View Time for each value of the variable Answer (i.e. if the participant successfully diagnosed the error, Answer = 1, if not, Answer = 0). Results suggest that successful detection of error, including models without an error, is linked with lower total time spent on each stimulus ($B$ = -0.3934, p <.0001). Furthermore, successful detection of semantic errors shows a faster time to first fixation on the area of interest (i.e. the zone containing the error) ($B$ = -0.3333, p = .0027). However, there are no significant results linking the detection of syntactic or no errors with the time to first fixation.

Hypothesis 2, which stipulates that successful error detection will be linked with fewer fixations, but with a higher proportion of fixations on the error, is tested using a Poisson regression with mixed model and a two-tailed level of significance of Fixation Count on Answer. A significant relation is found between successfully detecting an error in a model and lower fixation count ($B$ = -0.4402, p <.0001). Moreover, greater proportions of fixation are allocated to the zone containing semantic errors ($B$ = 0.5448, p <.0001) and syntactic errors ($B$ = 0.9379, p <.0001). However, while correct diagnosis of semantic errors are linked with a decrease in the number of fixation in the areas of interest ($B$ = -0,04956, p = .0897), the opposite is found for the successful detection of syntactic errors, where more fixations on the AOIs are required ($B$ = 0,3654, p <.0001).

In order to test Hypothesis 3, we apply a linear regression with mixed model and a two-tailed level of significance of Fixation Durations on Answer, allowing us to identify a significant correlation between successful diagnostics and shorter fixation duration ($B$ = -0.373, p <.0001). However, fixations in the AOIs are longer for successful diagnosed semantic errors ($B$ = 0.1654, p = .0061) and syntactic errors ($B$ = 0.4436, p <.0001), which indicate that the participants, when successfully identifying the errors, spend more time on the errors, but less time on the rest of the stimuli.

## 5. Discussion and Conclusion

Our preliminary results suggest that H1, H2 and H3 are partially supported. Significant links between successfully detecting an error and a lesser amount of time spent on a stimulus (H1), and between an accurate diagnostic and shorter fixation duration are found (H3). Furthermore, H2, which proposed fewer fixations and a

greater proportion of fixations on the errors when successfully detecting an error, is supported. However, successful detection of syntactic errors is significantly associated with a greater number of fixations in AOIs, which suggests that the error was detected, but the correction response is inhibited [9], possibly due to a higher level of complexity in syntactic errors. No significant link between syntactic errors and the number of fixations in the entire stimulus is found. Thus, this study presents evidence that the characteristics of visual attention of experienced modelers, such as lower number and duration of fixations, are generally related with successful error detection. On the other hand, attributes normally associated with novices, such as higher time spent on a stimulus or higher fixation duration, lead to unsuccessful error detection.

Our research so far offers both theoretical and practical contributions. The differences between the process and repertoire of attentional characteristics in the detection of semantic errors and syntactic errors reinforce Moody's [8] propositions about their complex interdependence. Syntactic errors require more attentional fixation than semantic errors. This finding runs contrary to our expectations and highlights the need for further studies to more fully articulate the differences and the metrics that might be used to measure them. The next phase of our research will address this challenge. From a practical standpoint, deeper insights into differences between the attentional characteristics will offer guidance to the evolution of BPMN and other notations and recommendations for curriculum development and training methods.

Limitations of this exploratory phase of our study center on the models used as stimuli for our experiment. While we tried to minimize the impact of domain-specific knowledge by using processes well-known to all potential participants, it is virtually impossible to negate the influence of the variation of domain familiarity between participants. Another limitation can be found in our sample. A bigger and more equally distributed sample will allow more complex statistical analyses and provide more significant findings. Finally, no measure was taken to evaluate the 'stopping rule', which is when a subject decides to terminate his information search because he judges that he has enough information to complete his task [24,25]. The next step of our research should evaluate this concept in order to better understand our eye tracking data, especially the measures linked to the view time.

## References


1. Recker, J.C., Dreiling, A.: Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In: Toleman, Mark, Cater-Steel, Aileen, & Roberts, Dave (Eds.) 18th Australasian Conference on Information Systems, , Toowoomba, Australia, pp. 356-366 (2007)
2. Parsons, J., Cole, L.: What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. Data & Knowledge Engineering **55**(3), 327-342 (2005)
3. Gemino, A., Wand, Y.: Comparing mandatory and optional properties in conceptual data modeling. In: Proceedings of the Tenth Annual Workshop on Information Technologies and Systems, Brisbane 2000, pp. 97-102 (2000)
4. Gemino, A., Wand, Y.: Evaluating modeling techniques based on models of learning. Communications of the ACM **46**(10), 79-84 (2003)
5. Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review **23**(4), 523-552 (2011)
6. Shanks, G.: Conceptual data modelling: an empirical study of expert and novice data modellers. Australasian Journal of Information Systems **4**(2) (1997)
7. Davis, C.J., Hevner, A.R., Labonte-LeMoyne, É., Léger, P.-M.: Expertise as a Mediating Factor in Conceptual Modeling. In: Information Systems and Neuroscience. pp. 85-92. Springer, (2018)
8. Moody, D.: The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on Software Engineering **35**(6), 756-779 (2009)
9. Van Waes, L., Leijten, M., Quinlan, T.: Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. Reading and Writing **23**(7), 803-834 (2009)
10. Henderson, J.M., Hollingworth, A.: The role of fixation position in detecting scene changes across saccades. Psychological Science **10**(5), 438-443 (1999).
11. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: A comprehensive guide to methods and measures. OUP Oxford, (2011)
12. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics **24**(6), 631-645 (1999)
13. Yusuf, S., Kagdi, H., Maletic, J.I.: Assessing the comprehension of UML class diagrams via eye tracking. In: Program Comprehension, 2007. ICPC'07. 15th IEEE International Conference, pp. 113-122. IEEE (2007)

14. Krupinski, E.A.: The importance of perception research in medical imaging. Radiation medicine **18**(6), 329-334 (2000)
15. Reingold, E.M., Sheridan, H.: Eye movements and visual expertise in chess and medicine. Oxford handbook on eye movements, 528-550 (2011)
16. Sheridan, H., Reingold, E.M.: Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. Front Psychol **5**, 941 (2014)
17. Glaser, R.: Education and thinking: The role of knowledge. American psychologist **39**(2), 93 (1984)
18. Lurigio, A.J., Carroll, J.S.: Probation officers' schemata of offenders: Content, development, and impact on treatment decisions. Journal of Personality and Social Psychology **48**(5), 1112 (1985)
19. Gemino, A., Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering **9**(4), 248-260 (2004)
20. Birkmeier, D., Kloeckner, S., Overhage, S.: An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users. In: ECIS, p. 2 (2010)
21. Bavota, G., Gravino, C., Oliveto, R., De Lucia, A., Tortora, G., Genero, M., Cruz-Lemus, J.A.: Identifying the weaknesses of UML class diagrams during data model comprehension. In: International Conference on Model Driven Engineering Languages and Systems, pp. 168-182. Springer (2011)
22. Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. Cognitive psychology **8**(4), 441-480 (1976)
23. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological bulletin **124**(3), 372 (1998)
24. Browne, G.J., Pitts, M.G.: Stopping rule use during information search in design problems. Organizational Behavior and Human Decision Processes **95**(2), 208-224 (2004)
25. Nickles, K.R., Curley, S.P., Benson, P.G.: Judgment-based and reasonging-based stopping rules in decision making under uncertainty, vol. 7285. University of Minnesota, (1995)

# Paying Attention Doesn't Always Pay Off: The Effects of High Attention Load on Evaluations of Ideas

Goran Calic[1], Nour El Shamy[1], Khaled Hassanein[1], and Scott Watter[2]

[1] DeGroote School of Business, McMaster University, Hamilton, Canada

{calicg,elshamyn,hassank}@mcmaster.ca

[2] Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Canada

watter@mcmaster.ca

**Abstract.** Creativity is a key driver of success for organizations in the digital age. Managers engaged in evaluating the creativity of new ideas are often subject to a myriad of technology-mediated distractors that compete for their attention. In this work in progress paper, we investigate whether attentional overload results in an upward bias for IT-mediated creativity evaluations. We report on promising early results that examines this phenomenon and set out to study its implications on IT design complexity.

**Keywords:** Creativity · Cognitive Load · Cognition · Electroencephalography · Eye Tracking · Pupil Dilation

## 1 Introduction

Creative ideas – ideas that are original and useful – are considered by most CEO's to be a crucial factor for success and sustainability in an ever turbulent environment [1]. Conversely, conventional ideas are discarded because they provide no competitive edge, as profits from these ideas have already been competed away [2]. Information Technology (IT) platforms are frequently utilized to facilitate creative idea generation and exchange between individuals. To this end, managers are investing in technology that fosters employee creativity [3, 4]. The objective of such investments is to get an edge in an increasingly competitive marketplace.

Previous Information Systems (IS) research has investigated how IT could be designed to foster creativity through priming – the presentation of a stimulus designed to subconsciously implant a concept in working memory that alters subsequent behavior [5]. Other studies have investigated the influence of IT design artefacts on users' cognitive demands during idea generation [6–8]. Much less is known about the idea evaluation process – the process of distinguishing creative ideas from conventional

ones. Particularly, it would be valuable to understand whether and how attentional load imposed by IT design artefacts influence users' ability to accurately evaluate the creativity of ideas.

When resources required to implement ideas are scarce, trial-and-error strategies of testing whether ideas are creative, rather than conventional, become prohibitively costly. Under such circumstances, a manager's capacity to make unbiased (i.e., accurate) evaluations about the creativity of an idea is highly valuable. Such evaluations require cognitive work. According to cognitive load theory [9], more attentional resources will result in a deeper understanding of the problem and its solution. In creative problems solving, this process begins with the retrieval of information from long-term memory and the subsequent analysis of that information in working memory. Attentional work that competes with this process may disrupt idea evaluation. Managers are increasingly engaged in IT-mediated work and online communication that competes for their attention. When presented with an idea online or through an Electronic Brainstorming (EBS) platform, manager evaluation may be biased due to the cognitive demands imposed by the system (e.g., system complexity) [10] or the attention-grabbing nature of other technology-mediated distractors (e.g., a new e-mail arrives, tweets, or even their cell phone) [11].

In this work in progress, we investigate whether attention overload biases IT-mediated creativity evaluations. We set out to test a broad conjecture that attention load will result in upward biased creativity evaluations. This is because attention load will attenuate an individual's capacity to make accurate judgments about an idea's future utility making the idea appear more surprising. A feeling of surprise has been shown to be highly related to the feeling of creativity and may result in the individual evaluating an idea as creative, even when it is conventional [12]. The implication is that in cognitive states of attention load, individuals will incorrectly evaluate conventional ideas as being creative. Evaluations of ideas are biased because they differ from evaluations the same individual would make in a state free of attention load. Thus, we coined the term "surprise bias" to describe this phenomenon.

## 2 Theoretical Background

In the next subsections we describe the theoretical lens used to develop our hypothesis about the link between attention overload and idea evaluation. Following this, we develop the concept of creativity and its relation to attention.

### 2.1 Attention Load

Cognitive load is the mental effort exerted by an individual to solve a problem or accomplish a task, during which information is retrieved from long term memory and temporarily stored in working memory for processing [13, 14]. While knowledge stored in long-term memory can be virtually unlimited, working memory capacity is limited to around seven elements or schemas of information at a time [14, 15]. In order to further integrate, compare, and process information in working memory,

additional working memory capacity is required, which further constrains the amount of information elements that can be simultaneously processed at a given moment [14].

Cognitive Load Theory (CLT) argues that task performance can be enhanced by optimizing cognitive resources through effective instructional design [9, 14]. CLT's design principles have been influential in understanding and enhancing individuals' IT-mediated creative ideation in Group Decision Systems (GSS) and Electronic Brainstorming (EBS). Potter and Balthazard [16] provide evidence that information load, in the form attending to others' feedback, reduced the number and quality of ideas generated by subjects in an EBS context. Heninger et al. [11] found that subjects who exchanged information simultaneously (i.e., contributing information while processing others' input) in GSS performed worse than subjects who performed these tasks separately. This difference was attributed to individual cognition and increased demands on working memory during dual-task performance rather than social processes [11]. Numerous studies building on CLT extend IS design principles that foster productivity by reducing cognitive demands [5, 8, 10, 17].

From a neurophysiological perspective, high attention load is associated with several electroencephalographic (EEG) and event-related potential (ERP) correlates as well as the dilation of the pupils. Attention load is associated with Event-Related Desynchronization (ERD) of neural oscillations in the Alpha band (8-12 Hz) [18, 19]. This effect can be observed in the higher Alpha band (10-12 Hz), particularly in the frontal region of the brain responsible for executive functions and working memory encoding [19, 20]. Additionally, under high attention load conditions, the amplitude of the P300 component during the presentation of stimuli is significantly reduced, indicating that cognitive resources are partly diverted away from the processing stimuli information [21, 22]. Further, attention load is associated with an involuntary dilation of the pupils, one of the sympathetic reactions of the Autonomous Nervous System (ANS) that prepare the organism to deal with stressful tasks [23].

### 2.2 Creativity

Creative ideas are defined as ideas that are both useful and unique in a particular context [24]. They are synthesized by reconstructing distant knowledge in novel ways to achieve a particular objective. Several EEG parameters have been found to correlate with creative ideation. Increases in the Alpha band power have been consistently observed during the generation of creative ideas [24–26]. Fink and Benedek [24] conclude that this relationship reflects an internal focused attention during the retrieval and recombination of distant knowledge from long-term memory. Additionally, the novelty and suddenness of a creative idea has been linked to increased activity in the Gamma band (38-44 Hz), particularly in the parieto-occipital region [27]. This has been interpreted as the result of successfully retrieving distant knowledge from long-term memory and a feeling of surprise, or "Aha!", by the creative idea [27].

Creativity is particularly valuable in situations characterized by attention load – an occurrence that has already pushed information processing of individuals to its limiting point. For example, a CEO might be burdened while evaluating the creativity of EBS-mediated responses to a rapidly escalating public relations crisis that can cost

millions of dollars. Research suggests that individuals suffering from attention load generate fewer and less creative ideas [16, 27–29]. Attention load will focus information processing to the task at hand and block the retrieval of remote and distant knowledge from long-term memory and knowledge which, when combined with task relevant easy to retrieve knowledge, results in the synthesis of creative ideas [30]. West [31] argues that competition, severity of challenge, time constraints imposed by the organization or environment, and other states which call on focused attention will inhibit creativity. Together, these literatures suggest that attention load inhibits the generation of creative ideas.

We suggest that attentional load also reduces evaluative capability, and in an unexpected way: attention load will bias managers to evaluate conventional ideas as creative. The implication is that such errors result in decision making errors – managers may invest in conventional ideas because they believe these ideas to be creative. During creative cognition, individuals rely on one or more heuristics (e.g., hypothesis testing [32], trial-and-error [33], or experiential search [34]) to form subjective beliefs about outcomes of implementing an idea. By reducing the information processing capacity available to form unbiased beliefs about the future utility of an idea, attention load increases an individual's subjective feeling of creativity, which we refer to as "surprise bias". Thus:

*Hypothesis: Higher attention load will result in higher creativity evaluations.*

## 3 Methodology and Preliminary Results

### 3.1 Experimental Design

An initial pilot study was conducted to test our hypothesis using EEG and eye tracking methods. We used Cognionics Quick-20 Dry EEG headsets to collect neural activity signals at 500 Hz. Tobii Pro X2-60 was used to collect eye tracking data including fixations and pupil diameter at 60 Hz. Noldus Syncbox was used to send markers to both Tobii Studio and Cognionics EEG Acquisition software to ensure proper synchronization between the systems.

The study followed a modified version of the Alternative Uses (AU) creativity experiment. AU tests are commonly utilized in the study of creative ideation using EEG, during which participants are asked to generate as many creative AU's for common objects [25, 26]. We modified this design by providing participants a pre-generated set of five AU's to each of three common day objects (i.e., brick, tin can, ping pong ball). We asked participants to evaluate each of the AU's provided on a scale of 1-5 (1- strongly disagree 5-strongly agree) on the two creativity dimensions of novelty and usefulness. E-Prime v 2.0 was utilized for the presentation of instructions and stimuli, and Noldus E-Prime Server was used to transfer E-Prime events and subject responses to the Observer XT software for analysis.

Participants were randomly assigned to either a baseline or an attention load condition. In the attention load treatment, participants were asked to simultaneously per-

form an auditory oddball task during the evaluation of AU's. Auditory stimuli, alternating between a frequent low-pitched tone and a rare high-pitched tone, were played every second for both conditions during the AU's tasks. Participants in the baseline condition were asked to disregard the noise, while those in the attention load condition were asked to count the number of times the target (i.e., high-pitched) tone was played and disregard the distractor tone. After completing the five AU's evaluation tasks for each object, participants in the attention load condition were asked to indicate the number of times the target tone was played. At the end of the experiment, participants in both conditions were asked to rate how distracting they felt the auditory stimuli were on a scale of 1-5 (1- not distracting at all 5-very distracting).

We received clearance from McMaster University's research ethics board, and 12 MBA students (i.e., 1 female, 11 male) were recruited through McMaster Digital Transformation Research Centre (MDTRC) and participated in the pilot study so far. Participants were screened for eye health related problems, and two participants were excluded from the EEG analysis due to handedness (i.e., left handed). The experiment lasted 20 minutes on average and participants received monetary compensation for their time.

### 3.2 Analysis and Preliminary Results

This research is still in its nascent stage, more participants are currently being recruited and data analysis is still underway. However, preliminary results seem to be promising. Participants in the attention load condition were more distracted ($M_{load}$ = 3.3) by the auditory stimuli than participants in the control group ($M_{control}$ = 2.4, $t$ = 1.8, $p$ = 0.055) indicating that the attention load manipulation is successful. Initial ERP analysis (Fig. 1) provides information on differences in cognitive processing between the two groups. Participants in the control group exhibited sustained voltage negativity over frontal, and especially prefrontal, electrode sites compared to relatively smaller negative values for the attention load condition. Extended negativity in the frontal region has been associated with greater executive control demands and focus on active task processing [35, 36]. In response to the onset of the AU trial, participants in the attention load condition are partly directing their cognition to process the auditory stimuli, and thus exhibit lower prefrontal negativity relative to the AU task [35, 36]. As hypothesized, creativity ratings are higher in the attention load than the baseline condition for 11 of the 15 AU's. While the differences are not statistically significant, the sample size at the time of submitting this work is too small to be conclusive. A larger sample to be tested in the near future will furnish a more rigorous test of the central hypothesis of this study.

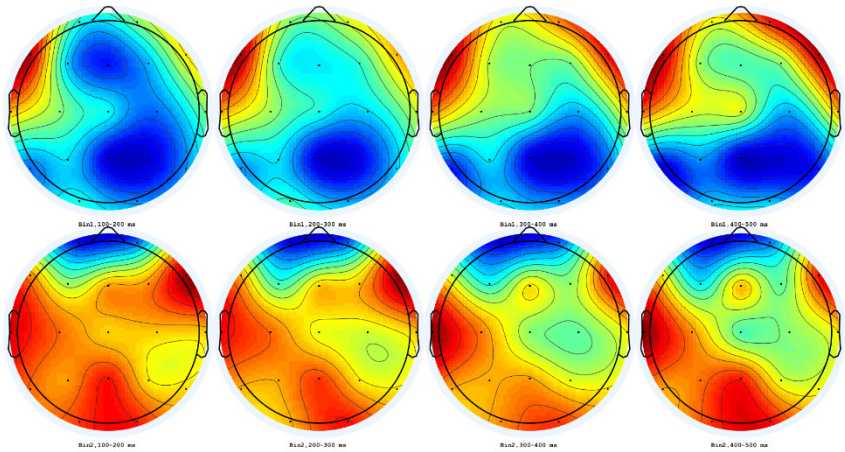



**Fig. 1.** Mean scalp voltage topography from 100ms to 500ms after the onset of AU's in 100ms intervals for the attention overload (top row) and control (bottom row) conditions.

## 4 Discussion and Next Steps

Our initial results are promising, yet analysis of the pupil dilation differences and EEG data is still ongoing, and more participants are required for these results to be conclusive. Prefrontal Alpha and parieto-occipital Gamma differences may particularly reveal interesting variations in executive control and evaluations of creative ideas under attention load. Our next steps will involve testing our hypothesis for IT-inflicted cognitive load in creativity evaluations. Particularly, we plan to study the impact of system complexity on managers' evaluation of EBS-mediated ideas, as well as on consumers' evaluation of crowdfunding ideas (e.g., Kickstarter). Studies have shown that ideas presented with more information and content had higher chances of success in attracting crowdfunding [37]. However, this goes against basic CLT design principles of parsimony in presenting information to avoid attention load [10]. It will be extremely rewarding to theory and practice to examine this phenomenon in our future work.

## References

1. IBM. 2010. IBM 2010 Global CEO Study: Creativity Selected as Most Crucial Factor for Future Success. *News releases*.
2. Gavetti, Giovanni. 2012. Toward a behavioral theory of strategy. *Organization Science* 23: 267–285.

3. Bhagwatwar, Akshay, Anne Massey, and Alan R. Dennis. 2013. Creative virtual environments: Effect of supraliminal priming on team brainstorming. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, 215–224. doi:10.1109/HICSS.2013.152.
4. Di Gangi, Paul M, Molly M Wasko, and Robert E Hooker. 2010. Getting Customers' Ideas To Work For You: Learning From Dell How To Succeed With Online User Innovation Communities. *MIS Quarterly Executive* 9.
5. Dennis, Alan R., Randall K. Minas, and Akshay P. Bhagwatwar. 2013. Sparking Creativity: Improving Electronic Brainstorming with Individual Cognitive Priming. *Journal of Management Information Systems* 29. IEEE: 195–216. doi:10.2753/MIS0742-1222290407.
6. Davern, Michael, Theresa Shaft, and Don Te'eni. 2012. Cognition Matters: Enduring Questions in Cognitive IS Research. *Journal of the Association for Information Systems* 13: 273–314.
7. Antunes, Pedro, and António Ferreira. 2011. Developing collaboration awareness support from a cognitive perspective. *Proceedings of the Annual Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2011.161.
8. Kolfschoten, Gwendolyn L. 2011. Cognitive load in collaboration - Brainstorming. *Proceedings of the Annual Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2011.107.
9. Sweller, John. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4: 295–312. doi:10.1016/0959-4752(94)90003-5.
10. Wang, Qiuzhen, Sa Yang, Manlu Liu, Zike Cao, and Qingguo Ma. 2014. An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems* 62. Elsevier B.V.: 1–10. doi:10.1016/j.dss.2014.02.007.
11. Heninger, William G, Alan R Dennis, and Kelly Mcnamara Hilmer. 2006. Research Note: Individual Cognition and Dual-Task Inference In Group Support Systems. *Information Systems Research* 17: 415–424. doi:10.1287/isre.l060.0102.
12. Simonton, Dean Keith. 2012. Creative Thought as Blind Variation and Selective Retention: Why Creativity is Inversely Related to Sightedness. *Journal of Theoretical and Philosophical Psychology* 33: 253–266. doi:10.1037/a0030705.
13. Sweller, John. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12: 257–285.
14. Sweller, John, Jeroen J . G . van Merrienboer, and Fred G . W . C . Paas. 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 10: 251–296.
15. Miller, George. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review* 63: 81–97.
16. Potter, Richard E., and Pierre Balthazard. 2004. The role of individual

memory and attention processes during electronic brainstorming. *MIS Quarterly* 28: 621–643.

17. Williams, Michael L., Alan R. Dennis, Antonie Stam, and Jay E. Aronson. 2007. The impact of DSS use and information load on errors and decision quality. *European Journal of Operational Research* 176: 468–481. doi:10.1016/j.ejor.2005.06.064.
18. Stipacek, A, R H Grabner, C Neuper, A Fink, and A C Neubauer. 2003. Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neuroscience Letters* 353: 193–196. doi:10.1016/j.neulet.2003.09.044.
19. Krause, Christina M., Lauri Sillanmäki, Mika Koivisto, Carina Saarela, Anna Häggqvist, Matti Laine, and Heikki Hämäläinen. 2000. The effects of memory load on event-related EEG desynchronization and synchronization. *Clinical Neurophysiology* 111: 2071–2078. doi:10.1016/S1388-2457(00)00429-6.
20. Klimesch, W, H Schimke, and G Pfurtscheller. 1993. Alpha Frequency , Cognitive Load and Memory Performance. *Brain Topography* 5: 241–251.
21. Parasuraman, Raja. 2011. Neuroergonomics: Brain, Cognition, and Performance at Work. *Current Directions in Psychological Science* 20: 181–186.
22. Lavric, Aureliu, Simon Forstmeier, and Gina Rippon. 2000. Differences in working memory involvement in analytical and creative tasks: an ERP study. *Cognitive Neuroscience* 11: 1613–1618. doi:10.1097/00001756-200006050-00004.
23. Riedl, René, and Pierre-Majorique Léger. 2016. *Fundamentals of NeuroIS*. Studies in Neuroscience, Psychology and Behavioral Economics. Berlin, Heidelberg: Springer.
24. Fink, Andreas, and Mathias Benedek. 2014. EEG alpha power and creative ideation. *Neuroscience and Biobehavioral Reviews* 44. Elsevier Ltd: 111–123. doi:10.1016/j.neubiorev.2012.12.002.
25. Fink, Andreas, Mathias Benedek, Roland H. Grabner, Beate Staudt, and Aljoscha C. Neubauer. 2007. Creativity meets neuroscience: Experimental tasks for the neuroscientific study of creative thinking. *Methods* 42: 68–76. doi:10.1016/j.ymeth.2006.12.001.
26. Dietrich, Arne, and Riam Kanso. 2010. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin* 136: 822–848. doi:10.1037/a0019749.
27. Sandkühler, Simone, and Joydeep Bhattacharya. 2008. Deconstructing Insight: EEG Correlates of Insightful Problem Solving. Edited by Paul Zak. *PLoS ONE* 3: e1459. doi:10.1371/journal.pone.0001459.
28. de Dreu, Carsten K.W., Bernard A. Nijstad, Matthijs Baas, Inge Wolsink, and Marieke Roskes. 2012. Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin* 38: 656–669. doi:10.1177/0146167211435795.

29. Smith, Steven M. 1995. Getting into and out of mental ruts: A theory of fixation, incubation, and insight. The MIT Press.
30. Kasof, Joseph. 1997. Creativity and Breadth of Attention. *Creativity Research Journal* 10: 303–315. doi:10.1207/s15326934crj1004_2.
31. West, Michael A. 2002. Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied psychology* 51: 355–387.
32. Finke, Ronald A, Steven M Smith, and Thomas B Ward. 1996. *Creative cognition theory, research, and applications*. Cambridge, Massachusetts: MIT Press.
33. Campbell, Donald T. 1960. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review* 67: 380–400.
34. Getzels, Jacob W. 1987. Creativity, intelligence, and problem finding: Retrospect and prospect. *Frontiers of creativity research*: 88–102.
35. Watter, Scott, Jennifer J. Heisz, James W. Karle, Judith M. Shedden, and Ivan Kiss. 2010. Modality-specific control processes in verbal versus spatial working memory. *Brain Research* 1347: 90–103. doi:10.1016/j.brainres.2010.05.085.
36. Kiss, Ivan, Scott Watter, Jennifer J. Heisz, and Judith M. Shedden. 2007. Control processes in verbal working memory: An event-related potential study. *Brain Research* 1172: 67–81. doi:10.1016/j.brainres.2007.06.083.
37. Du, Quianzhou, Zhilei Qiao, Weiguo Fan, Mi Zhou, Xuan Zhang, and Alan G. Wang. 2015. Money Talks: A Predictive Model on Crowdfunding Success Using Project Description. In *ACIS 2015*, 1–8.

# Using Gaze Behavior to Measure Cognitive Load

Lisa Perkhofer[1], and Othmar Lehner[2]

[1] University of Applied Sciences Upper Austria, Management Accounting, Steyr, Austria
[2] Said Business School, University of Oxford, Oxford, UK
lisa.perkhofer@fh-steyr.at; othmar.lehner@sbs.ox.ac.uk

**Abstract.** Measuring and influencing cognitive load during information processing can be seen as a promising instrument to mitigate the risk of information overload while increasing processing capabilities. In this study, we demonstrate how to use cross-sectional time-series data generated with an eye tracking device to indicate cognitive load levels. Thereby we combine multiple measures related to fixations, saccades and blinks and calculate one comprehensive and robust measure. Replicability and applicability is demonstrated by conducting two separate experiments in a decision-making scenario in the context of information visualization.

**Keywords:** eye tracking; cognitive load; structural equation modelling

## 1 Introduction

Cognitive load is defined as "the amount of working memory resources required in cognitive task execution" [1] p. 381. The monitoring of cognitive load levels during information processing and learning in situations in which a user operates near his or her working memory capacity limits can be identified as major application [2]. If these capacity limits are breached the user enters the state of information overload in which processing capabilities are seriously impaired, further learning is prohibited, and thus biased decisions may often be the consequence [3, 4]. To avoid these disadvantageous effects, research in the field of human-computer interaction, with topics such as interface design and visual computing has begun to focus on the construct of cognitive load already at design and evaluation stages [1, 5–13].

Consequently, multiple models for the assessment of cognitive load have been suggested so far [1, 7, 14]. Proposed measurement possibilities range from simple self-assessment to a more high-end analysis based on fMRI [14]. Unfortunately, when it comes to the actual application in empirical studies, physiological measures are still scarce especially when it comes to information visualization [9]. Based on the foundational paper by Zagermann et al. [12], this study therefore contributes to closing this gap by empirically assessing various complementary eye tracking measures as proxies for a single comprehensive and robust cognitive load measure. Two separate experiments in the context of InfoVis were conducted and analyzed to demonstrate reliability of the construct. Furthermore, its applicability is tested in a real-case decision-making scenario and evaluated based on a structural equation model.

## 2 Cognitive load measurement

Cognitive load is a construct, which is not directly observable, but can be assessed using indicators [14]. Brücken et al. [14] summarize different measurement models for those indicators, splitting in subjective measurement via self-reports and objective measurements via physiological, behavioral, outcome-oriented or brain activity measures. Up to now however, the majority of studies conducted in InfoVis measure cognitive load either by analyzing results on decision-making outcome (task accuracy or task completion time), or by subjective assessment via self-reported data (e.g. NASA TLX, SWAT) [9, 15, 16]. This contemporary practice misses robustness and is therefore susceptible to multiple measurement biases, possibly even leading to spurious correlations [16]. Further results cannot be used reliably for prediction or real time assessment given an optimal and user specific support. More robust methods from NeuroIS could help in triangulating early results from literature and further produce more robust empirical evidence which can also be used real time [9].

One physiological measurement method, which is increasingly gaining attention in this context, is eye tracking [13, 16–19]. By relying on eye tracking related measures it should be possible to distinguish between the mental demand of different interface or system designs [7, 8] while predicting it's influence on decision-making outcome during cognitive task execution [6, 13]. In a context of static visual stimuli, eye movements measures on fixations and saccades (mostly voluntary), but also measures on pupil dilation and blink related data (mostly involuntarily movement), have been associated with an increase in mental demand [18]. In the following, we are going to individually and shortly describe possible eye movement indicators and summarize their respective implications in the context of cognitive load.

- **Fixations.** Fixations are cognitively controlled short dwells where the eye stops, and one processes information. An increase in fixation duration is said to indicate higher demands on working memory, whereas a high fixation rate signals great visual and/or cognitive complexity (task dependent searching behavior) [17–19].
- **Saccades.** Saccades refer to the shifts between fixation locations. During the actual movement of the eye, no information acquisition can take place because one is blind. However, it has been shown that a long saccade length indicates higher cognitive load [18] and also a high saccade count is associated with higher visual complexity [13].
- **Pupil.** Most studies on cognitive load and eye tracking focus on pupilometry [13]. In states of high mental load, the pupil diameter enlarges proportionally [1, 13, 20]. However, in order to account for individual and situational differences it is necessary to evaluate pupil diameter with respect to an adaptive baseline [13].
- **Blinks.** Blinking is said to indicate information processing and they occur before and after high states of cognitive load. High blink rates as well as long blink durations are associated with higher mental demand and thinking [18, 20].

## 3 Development of one comprehensive construct

Looking at the data generated by eye tracking devices, we are confronted with cross-sectional time-series data [21]. Each task results in a scan-path, which is a series of fixations, saccades and blinks stringed together. More precisely, a scan-path starts with the onset of the stimuli and ends with its offset and it further includes all periods of task dependent visual search behavior as well as all periods of active mental processing [17, 18]. What we can observe is therefore not a linear trend but a tendency to mean reversion for all eye tracking measures (an example for mean reversion of saccade duration, and pupil diameter is presented in Figure 1). This means in case of high mental load, which according to previous research indicates mental processing, high states of saccade duration, pupil diameter, blink duration etc. can be observed [16, 20].

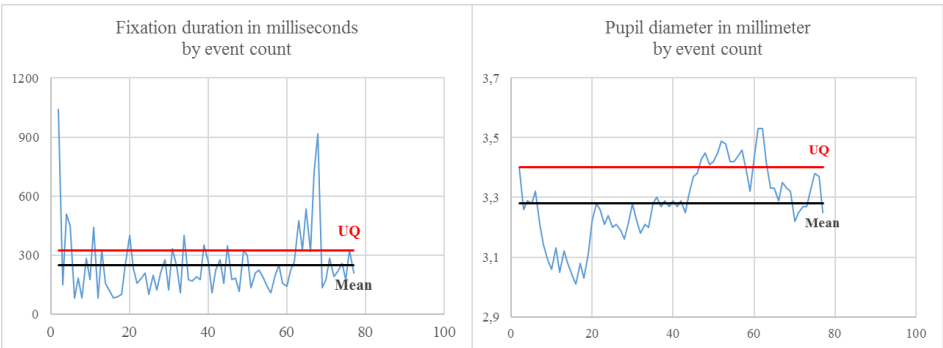



Figure 1: Mean reversion tendency of fixation duration and pupil diameter (example one participant, one stimulus and one task)

Summing up the findings as presented in the previous section and shown in Figure 1, we can see an increase in value in all measures in stages of actual mental processing, while during visual search periods values should normalize and return to average. To compare the mental demand of one stimulus with another, comparing their respective peaks seems logical, as these should represent actual processing periods. To mitigate the effect of outliers but still focus on phases of high mental demand, we only focus on the upper quintile (red line in Figure 1) of each eye tracking related measure (saccade duration, fixation duration, blink duration, pupil diameter). For aggregation of the resulting time-series a quantile-regression form (weighted average) is applied [21].

This procedure allows for a calculation of individual values per task, stimulus and participant for all relevant eye tracking measures, which can be used for further assessment. In contrast to most studies, we are trying to create one comprehensive measure (including information on fixations, saccades, pupil dilation, and blinks) following the example of Siegle et al. [20] instead of focusing on just one. Eye tracking indicators provide complementary information but their combination into one comprehensive construct necessitates the use of partial least squares structural equation modelling.

## 4 Experiments

As indicated in section 2, eye tracking related measures used in this experiment were saccade duration, saccade count, fixation duration, fixation count, pupil diameter, pupil diameter difference, blink duration, and blink count. While count related measures were used as is, duration related measures were calculated according to the procedure explained in section 3. With respect to the baseline introduced for pupil dilation we used the median of the first five fixations per task as after and before task execution pupil dilation is normalized and this allowed us to account for situational and personal differences [13, 22]. Furthermore, constant lightning conditions and an undisturbed and quiet environment were ensured throughout the experiment.

- In **study I** 118 students participated in the experiment. Participants were randomly assigned to one out of four groups. They had to answer questions with varying complexity levels (4 in total) based on different visualization types (16 in total). 1,888 observations were recorded for analysis[1].
- In **study II** 60 students participated in the experiment. Participants were again randomly assigned to one out of three groups. They had to answer questions with varying complexity levels (3 in total) based on visualization dashboards with three levels of data density (3 in total). 720 observations were recorded for analysis[1].

Data was recorded using an SMI stationary eye tracking system with nine-point calibration and 120 Hz sampling frequency. Eye tracking data was recorded and analyzed with Experiment Center as well as BeGaze (Version 3.7). The stimulus material was presented in randomized order in each experiment. For the experimental tasks in both studies, participants slipped into the role of the CEO of a fictitious company.

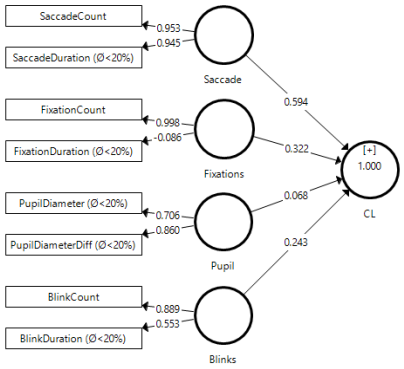
Eye tracking measures were only used if the quality of recordings was high: The tracking ratio per stimulus, which is the time being recorded by the eye tracking system divided by the time of the stimulus, needed to be above 95%. In the case of missing or excluded values mean replacement was used. For each first order construct (saccade, fixations, pupil, and blinks), two indicators were measured. For CL a repeated measurement approach was used, meaning that each indicator used for the first order construct is reused for the second order construct representing CL. This approach is the most frequently used method for estimating higher-order constructs and allows to assess the impact of each eye tracking related event in order to gain a more compressive understanding of cognitive load. As a result predictability on decision-making outcome per user can be evaluated. For the generation of one comprehensive measure we used PLS-SEM (SmartPLS version 3.2.7) [23].

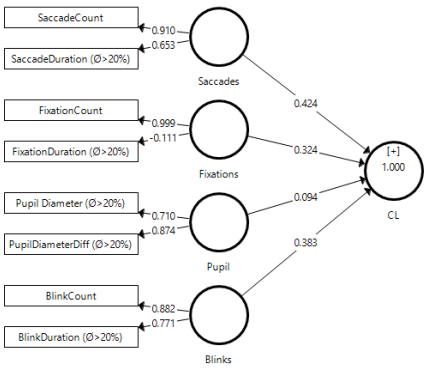
[1] Task types and the stimulus material can be downloaded from the author's homepage.

**Table 1.** Results on Study I and Study II

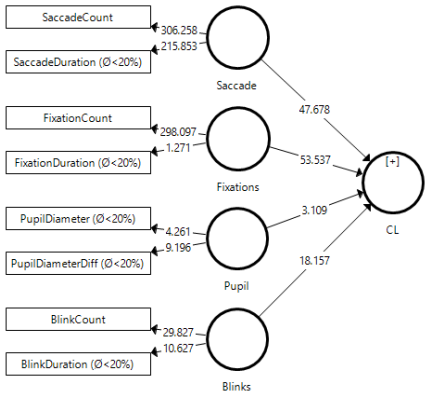
| Study I | Study II |
|---|---|

**Step 1: Evaluation of the second order formative construct**

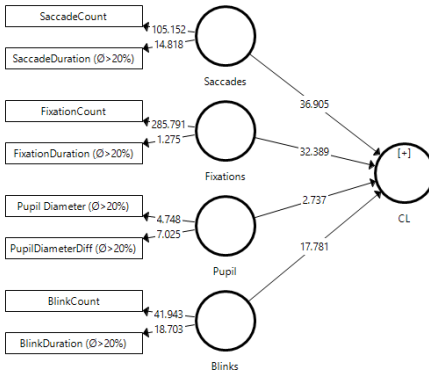



Bootstrapping routine (t-statistics)




Bootstrapping routine (t-statistics)




Cross loadings:

| | Blinks | Fixations | Pupil | Saccade | CL |
|---|---|---|---|---|---|
| BlinkCount | **0.889** | 0.543 | 0.009 | 0.299 | 0.569 |
| BlinkDuration (Ø<20%) | **0.553** | 0.181 | -0.004 | 0.202 | 0.313 |
| FixationCount | 0.545 | **0.998** | 0.068 | 0.740 | 0.898 |
| FixationDuration (Ø<20%) | 0.062 | **-0.086** | -0.078 | -0.074 | -0.062 |
| PupilDiameter (Ø<20%) | -0.025 | 0.043 | **0.706** | 0.074 | 0.099 |
| PupilDiameterDiff (Ø<20%) | 0.026 | 0.069 | **0.860** | 0.086 | 0.138 |
| SaccadeCount | 0.355 | 0.776 | 0.091 | **0.953** | 0.909 |
| SaccadeDuration (Ø<20%) | 0.294 | 0.628 | 0.103 | **0.945** | 0.842 |

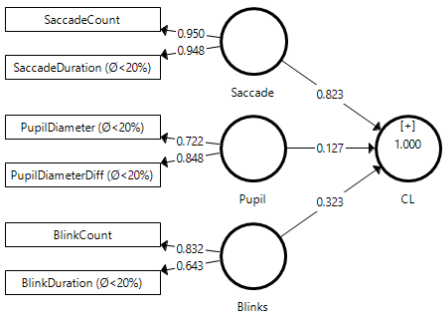
Cross loadings

| | Blinks | Fixations | Pupil | Saccades | CL |
|---|---|---|---|---|---|
| BlinkCount | **0.882** | 0.438 | 0.035 | 0.569 | 0.72 |
| BlinkDuration (Ø>20%) | **0.771** | 0.289 | 0.035 | 0.337 | 0.53 |
| FixationCount | 0.455 | 0.999 | 0.103 | 0.891 | 0.88 |
| FixationDuration (Ø>20%) | 0.102 | -0.111 | -0.059 | -0.085 | -0.03 |
| Pupil Diameter (Ø>20%) | -0.039 | 0.086 | **0.710** | 0.120 | 0.13 |
| PupilDiameterDiff (Ø>20%) | 0.084 | 0.084 | **0.874** | 0.112 | 0.18 |
| SaccadeCount | 0.524 | 0.989 | 0.116 | **0.910** | 0.91 |
| SaccadeDuration (Ø>20%) | 0.346 | 0.255 | 0.118 | **0.653** | 0.50 |

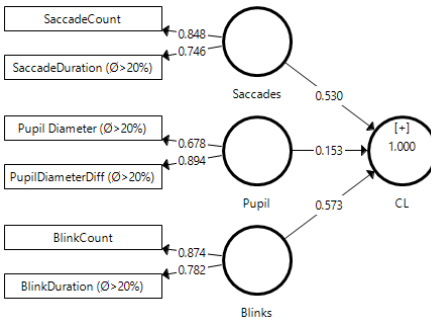
Construct reliability is good for all measures (above the 0.5 threshold for AVE), however, loadings on the construct fixation indicate a problem in both experimental studies. Fixation duration is not significant and has a negative impact on cognitive load. Additioanlly fixation count shows colliniarity with saccade count, blink count and saccade duration (highlightes in red). Based on these two observations, we decided to exclude the first oder construct fixation from the second order construct cognitive load.

**Step 2: Evaluation of the formativ construct without fixations**

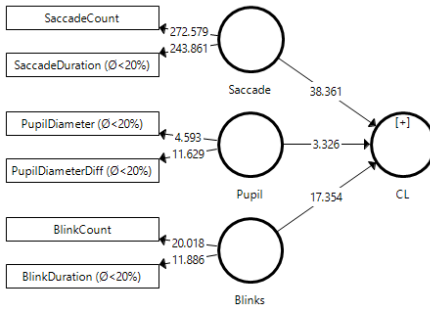
**Study I**



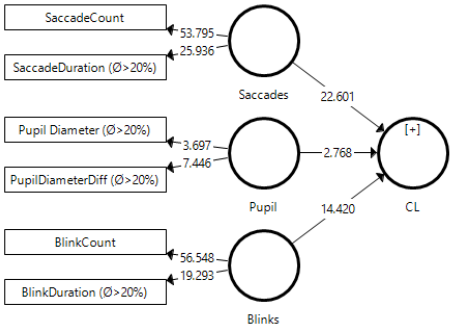

**Study II**




Bootstrapping routine (t-statistics)




Bootstrapping routine (t-statistics)




Cross loadings:

| | Blinks | Pupil | Saccade | CI |
|---|---|---|---|---|
| BlinkCount | **0.832** | 0.009 | 0.298 | 0.51 |
| BlinkDuration (Ø<20%) | **0.643** | -0.005 | 0.202 | 0.37 |
| PupilDiameter (Ø<20%) | -0.030 | **0.722** | 0.075 | 0.14 |
| PupilDiameterDiff (Ø<20%) | 0.028 | **0.848** | 0.086 | 0.18 |
| SaccadeCount | 0.356 | 0.091 | **0.950** | 0.90 |
| SaccadeDuration (Ø<20%) | 0.292 | 0.103 | **0.948** | 0.88 |

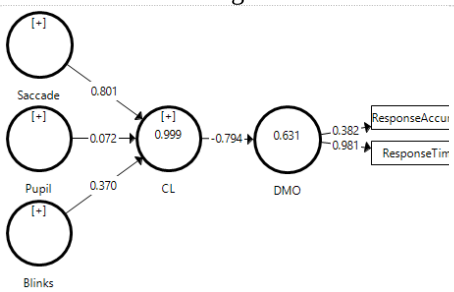
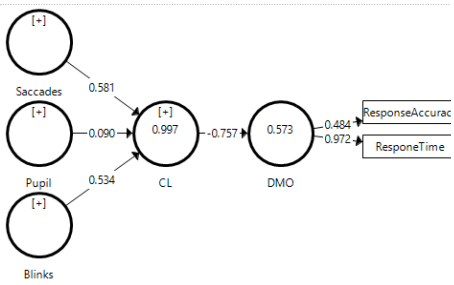
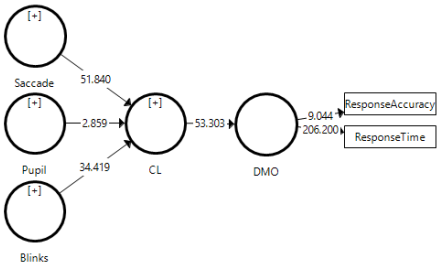
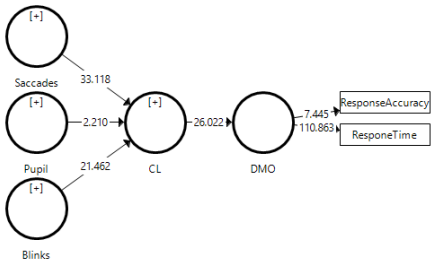
Cross loadings:

| | Blinks | Pupil | Saccades | CL |
|---|---|---|---|---|
| BlinkCount | **0.874** | 0.039 | 0.564 | 0.806 |
| BlinkDuration (Ø>20%) | **0.782** | 0.039 | 0.327 | 0.627 |
| Pupil Diameter (Ø>20%) | -0.039 | **0.678** | 0.127 | 0.149 |
| PupilDiameterDiff (Ø>20%) | 0.085 | **0.894** | 0.111 | 0.245 |
| SaccadeCount | 0.522 | 0.116 | **0.848** | 0.767 |
| SaccadeDuration (Ø>20%) | 0.343 | 0.115 | **0.746** | 0.610 |

Excluding fixation as a construct does also mean excluding the respective indicators form the CL measure. All three remaining first order constructs are reliable (AVE is above 0.5) and no colliniarity issues or cross loadings (except a small exceedance in the second experiment with respect too blink count and saccade count) in the outer model can be observed. By running a bootstrapping routine also significance for all measures ($p<0.01$) can be obtained, which is indicated by t-staticitcs (values above 1.96 indicate significance below a 0.05 level). With respect to the weights of the first order constructs we can see that saccadic and blink related data have a stronger explainatory power than pupil related information.

## Step 3: Using the formativ construct to predict decision making-outcome

In this last step, we show applicability of the introduced cognitive load measure in a decision-making context. The new construct has a high explainatory power on decision making outcome measured by task accuracy and task time ($R^2$ is roughly

| Study I | Study II |
|---|---|
| 60%) in both conducted studies. It can be shown, that the higher cognitive load, the lower decision making outcome. | |
|   |   |
| Bootstrapping routine (t-statistics) | Bootstrapping routine (t-statistics) |
|   |   |

## 5 Conclusion

As discussed in the beginning, measuring cognitive load is of high relevance for practice because system and interface designers need to avoid creating situations of information overload for the respective users [24, 25]. Therefore, the field is in need of a comprehensive and reliable measure beyond the old proxies of time and error [26]. In this study, we propose to include physiological measures based on eye tracking data to determine cognitive load levels, which then can be used to compare different tasks as well as complete system designs. Some early form of reliability of the proposed measurement is demonstrated as both experiments show similar results although different tasks and different visual stimuli were used. Further, its applicability in InfoVis is shown, because both studies were embedded in decision-making scenarios which is representative for the field and the explanatory power of cognitive load on decision making outcome was high in both ($R^2$ is roughly 60%). As potential future research endeavor, a more robust regression method may be applied to further improve accuracy, the authors suggest a sampling lasso quantile regression for this.

As potential implications, the introduced measure could further help discern differences between tools and design options, and enhance our understanding of individual differences [5]. By doing so, it could help researchers in quantifying personalization and adaption needs [1] and can even be used in-situ to change properties or layout features according to the respective user's cognitive state [18].

## References


1. Hossain, G.; Yeasin, M.: Understanding Effects of Cognitive Load from Pupillary Responses Using Hilbert Analytic Phase. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, pp. 381–386 (2014)
2. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. In: Learning and Instruction, vol. 4, no. 4, pp. 295–312 (1994)
3. Sweller, J.: Element interactivity and intrinsic, extraneous, and germane gognitive load. In: Educ Psychol Rev, vol. 22, no. 2, pp. 123–138 (2010)
4. Bawden, D.; Robinson, L.: The dark side of information: Overload, anxiety and other paradoxes and pathologies. In: Journal of Information Science, vol. 35, no. 2, pp. 180–191 (2009)
5. Huang, W.; Eades, P.; Hong, S. H.: Measuring effectiveness of graph visualizations: A cognitive load perspective. In: Information Visualization, vol. 8, no. 3, pp. 139–152 (2009)
6. Falschlunger, L.; Lehner, O.; Treiblmaier, H.: InfoVis: The Impact of Information Overload on Decision Making Outcome in High Complexity Settings. In: Proceedings of the 2016 SIG HCI, Dublin, pp. 1–5 (2016)
7. Anderson E. W.; Potter, K.C.; Matzen, L. E.; Shepherd, J. F.; Preston, G. A.; Silvia, C.T.: A user study of visualization effectiveness using EEG and cognitive load. In: Computer Graphics Forum, vol. 30, no. 3, pp. 791–800 (2011)
8. Cole, M. J.; Gwizdka, J.; Liu, C.; Belkin, N. J.; Zhang, X.: Inferring user knowledge level from eye movement patterns. In: Information Processing & Management, vol. 49, no. 5, pp. 1075–1091 (2013)
9. Dimoka, A.; Pavlou, P. A.; Davis, F. D.: Research Commentary —NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. In: Information Systems Research, vol. 22, no. 4, pp. 687–702 (2011)
10. Kalyuga, S.: Effects of learner prior knowledge and working memory limitations on multimedia learning. In: Procedia - Social and Behavioral Sciences, vol. 83, pp. 25–29 (2013)
11. Falschlunger, L.; Treiblmaier, H.; Lehner, O.: Cognitive differences and their impact on information processing: An empirical study comparing survey and eye tracking data. In: Information Systems and Neuroscience and Organizations, Gmunden, pp. 137–144 (2015)
12. Weber, B.; Neurauter, M.; Pinggera, J.; Zugal, S.; Furtner, M.; Martini, M.; Sachse, P.:

    Measuring Cognitive Load During Process Model Creation. In: Information Systems and Neuroscience and Organizations, Gmunden, pp. 129–136 (2015)

13. Lallé, S.; Conati, C.; Carenini, G.: Prediction of individual learning curves across information visualizations. In: User Model User-Adap Inter, vol. 26, no. 4, pp. 307–345 (2016)

14. Brücken, R.; Plass, J. L.; Leutner, D.: Direct measurement of cognitive load in multimedia learning. In: Educational Psychologist, vol. 38, no. 1, pp. 53–61 (2003)

15. Paas, F.; Tuovienen, J. E.; Tabbers, H.; Van Gerven, P. W. M.: Cognitive load measurement as a means to advance cognitive load theory. In: Educational Psychologist, vol. 38, no. 1, pp. 63–71 (2003)

16. Paas, F.; van Merriënboer, J. J. G.; Adam, J. J.: Measurement of cognitive load in instructional research. In: Perceptual and motor skills, vol. 79, no. 1 Pt 2, pp. 419–430 (1994)

17. Wang, Q.; Yang, S.; Liu, M.; Cao, Z.; Ma, Q.:An eye-tracking study of website complexity from cognitive load perspective. In: Decision Support Systems, vol. 62, pp. 1–10 (2014)

18. Zagermann, J.; Pfeil, U.; Reiterer, H.: Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. In: Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16, Baltimore, MD, USA, pp. 78–85 (2016)

19. Goldberg, J. H.; Helfman, J. I.: Comparing information graphics: A critical look at eye tracking. In: Proceedings of the 3rd BELIV'10 Workshop BEyond time and errors novel evaLuation methods for Information Visualization, pp. 71–79 (2010)

20. Siegle, G. J.; Ichikawa, N.; Steinhauer, S.: Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. In: Psychophysiology, vol. 45, no. 5, pp. 679–687 (2008)

21. Marcus, A.; Harding, M.; Lamarche, C.: Quantile regression for time-series-cross-section data. In: International Journal of Statistics and Management System, vol. 6, no. 1-2, pp. 47–72 (2011)

22. Iqbal, S. I.; Adamczyk, P. D.; Zheng, X. S.; Bailey, B. P.: Towards an indes of opportunity: Understanding changes in mental workload during task execution. In: Proceedings of the CHI 2005, Portland, Oregon, USA, April 2-7, pp. 311–320 (2005)

23. Hair, J.; Hollingsworth, C. L.; Randolph, A. B.; Chong, A. Y. L.: An updated and expanded assessment of PLS-SEM in information systems research. In: Industr Mngmnt & Data Systems, vol. 117, no. 3, pp. 442–458 (2017)

24. Miller, G. A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. In: Psychological Review, vol. 101, no. 2, pp. 343–352 (1956)

25. Baddeley, A. D.; Hitch, G. J.: Development of working memory: should the Pascual-Leone and the Baddeley and Hitch models be merged?. In: Journal of experimental child psychology, vol. 77, no. 2, pp. 128–137 (2000)

26. Elmqvist, N.; Yi, J. S.: Patterns for visualization evaluation. In: Proceedings of the 2012 BELIV Workshop on Beyond Time and Errors - Novel Evaluation Methods for Visualization - BELIV '12, Seattle, Washington, pp. 1–8 (2012)

# The impact of gestures on Formal Language learning and its neural correlates: A study proposal

Wriessnegger, Selina Christin[1]; Hacker, Christopher[1]; Macedonia, Manuela[2]; Müller-Putz, Gernot R.[1]

[1] Graz University of Technology, Graz, Austria
{s.wriessnegger, gernot.mueller}@tugraz.at
Christopher.hacker@students.tugraz.at
[2] Johannes Kepler University Linz, Austria
manuela@macedonia.at

**Abstract.** This pilot study reports about the impact of gestures on learning a formal language like Python. The aim of this research-in-progress is to find out if memory performance will benefit from the coupling of gestures and words in the learning phase. Previous research has demonstrated that gestures accompanying speech have an impact on memory for verbal information. This is the first study applying the body-mind concept to formal language learning. We introduce the study design and the results of one person.

**Keywords:** Formal language, Memory, Embodiment, Gestures, EEG

## 1 Introduction

Previous studies on language acquisition have demonstrated that performing representative gestures during encoding (enactment) enhances memory for concrete words. Furthermore, gestures support memory for verbal information in native [1] and foreign language [2]. There is neuroscientific evidence for the fact that words are not abstract symbols [3], moreover, they are at the basis of the body [4,5,6,7,8].
Macedonia and Klimesch [9] showed that learning novel words with gestures enhances memory compared to learning novel words audio-visually in the long-term range. They suggested that words learned audio-visually are shallow and decay fast. In contrast, complex sensorimotor codes created by pairing a novel word with a gesture are deep, that is retaining information better and decay slower. For instance, if the word is connected to a motor act that occurred during learning, the word's network comprises a motor component. Another study showed that hearing “pick” activates the cortical region that controls hand movements, “kick” activates foot movements, and “lick” activates tongue movements [10].
The prediction that learners have better memory for words encoded with gestures has further been supported by [11,12]. They investigated the impact of enactment on

abstract word learning in a foreign language. In a transfer test, participants produced new sentences with the words they had acquired. Items encoded through gestures were used more frequently, demonstrating their enhanced accessibility in memory. The results are interpreted in terms of embodied cognition. The results of a recently performed EEG study with 30 participants, based on the work of Macedonia and colleagues [2,11,12], gives further evidence for the beneficial impact of learning with gestures on memory performance. Based on these findings, we were interested if the body-mind concept could also have a beneficial impact on formal language acquisition. Furthermore we will investigate the neural coupling of cognition and action by means of EEG. Since a formal language is quite similar to a natural language in terms of syntax and semantics, we hypothesized that the learning process will also be enhanced if it goes along with gestures. In this paper we introduce the preliminary experimental procedure, the study design and the results of one test session.

## 2 Materials and Methods

### 2.1 Participants

One female student (21 years) participated in this pilot study. She had normal vision and no experience in programming. She gave written consent participating in the study.

### 2.2 Experimental Paradigm

Before the experiment started the experimenter carefully instructed the participant about the purpose and the procedure of the following pilot study. The introduction included the studying of a document which contains a detailed description of the programming code of a calculator (Fig. 1). Figure 1 shows an excerpt of the document (first 3 out of 13 functions) with the different functions and their explanations. The other functions include the choice of operation, the numbers the participants selected and the final output (see Fig.3, left).

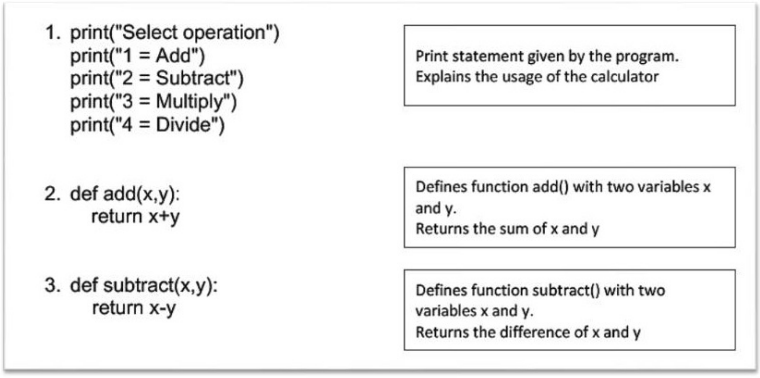



**Figure 1:** The first three functions and definitions of the calculator

In total 13 code parts with a detailed description of the functions were shown. The subject was additionally advised to the importance of the sequence of code lines. The task was to learn the different functions of the formal language (Python) resulting in a small calculator program. The functions were shown by short videos. Each variable was assigned to one specific gesture. The gestures were symbolically meaningful to the variables and operations they present. For example the first picture in Fig.2 shows a part of the movie representing the variable "`def`". The subject had to learn the functions with the according gestures 3 times on 3 different days (Fig. 2).

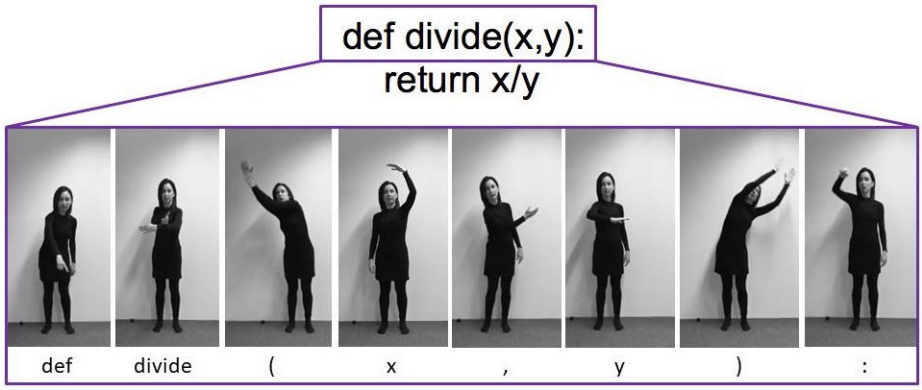



**Fig. 2.** Example of a function with corresponding gestures.

Fig.2 shows the sequence of the gestures which are assigned to the function "`def divide(x,y):`", defining the function divide() with two variables x and y. After each learning phase the participant had to fill out a google questionnaire to control the learning progress. The whole program consisted of 13 functions including 42 gestures which have been imitated by the subject. The training schedule included 3 training

days with a repetition of the 13 functions 2 times on each day (Table 1). Participants learned the different variables, resulting in 13 corresponding functions, by watching and imitating the gestures they saw and the video.

| **DAY 1** | **DAY 2** | **DAY 3** |
|---|---|---|
| Training 1 | Training 2 | Training 3 |
| Session 1 (20 min)<br>Session 2 (20 min) | Session 1 (20 min)<br>Session 2 (20 min) | Session 1 (20 min)<br>Session 2 (20 min) |
| *Break (5 min)* | *Break (5 min)* | *Break (5 min)* |
| Recall (Google form) | Recall (Google form) | Coding Test |

**Table 1:** Training Schedule Pilot study

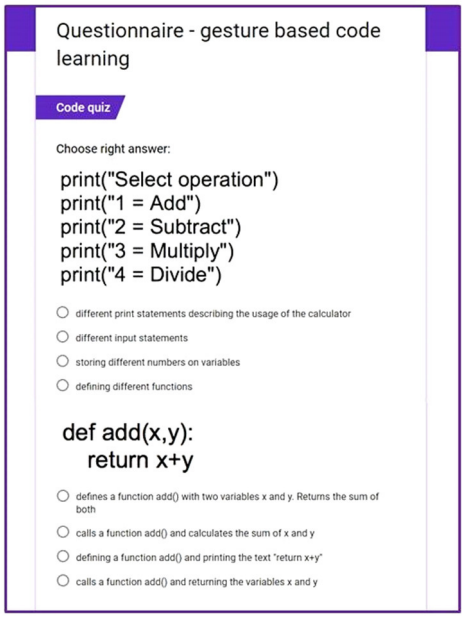



**Fig. 2.** Screenshot of the used questionnaire in the recall test

On the third day (session 3) the recall test consisted of the programming of the calculator, that is, the participants have to remember all 13 functions in the correct order. For this purpose, an editor was used where participants will get immediately feedback about the success of the recall phase (Fig.3).

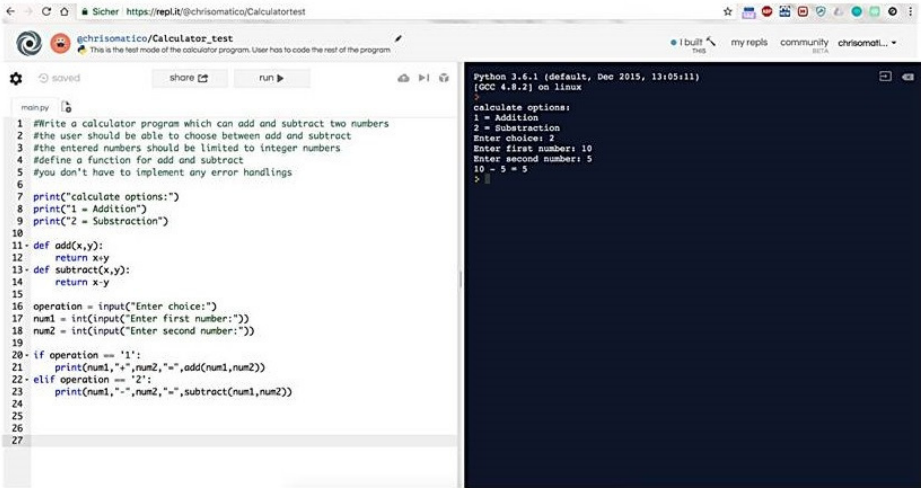



**Figure 3:** Screenshot of the editor in the final recall phase "programming the calculator".

## 2.3 EEG Recording

In the main experiment we will additionally record the EEG. For this purpose we will use a mobile and wireless 32 channel EEG system (LiveAmp;Brain Products). Beside the behavioral recall performance we are also interested in the brain activity during the recall phase (memory processes) and especially what does the brain do when the subject is moving, that is performing the gestures in the learning phase.


**Fig.4.** Mobile EEG System (LiveAmp Brain Products).

For example [12] found in their fMRI study, activity in the premotor cortices for words encoded with iconic gestures. In contrast, words encoded with meaningless gestures elicited a network associated with cognitive control. These findings suggest that memory performance for newly learned words is not driven by the motor component as such, but by the motor image that matches an underlying representation of the word's semantics. We were interested in finding comparable results by means of EEG. Moreover, by using a sophisticated connectivity analysis we will further investigate correlations between activity in motor related and frontal areas due to the body-mind concept.

To test the usability of the program and the study design we made a pilot study with one subject. In the following the results of this test session will be presented.

## 3 Results of the Test Session and Discussion

The student learned the different functions of the calculator on three different days. In the first and second recall test she made some errors like forgetting a semicolon, colon or bracket. Nevertheless, surprisingly she answered all questions regarding the functions of the different codes correctly. In the last recall phase, which required the programming of the whole calculator, she was also successful. The result is shown in Fig.5.

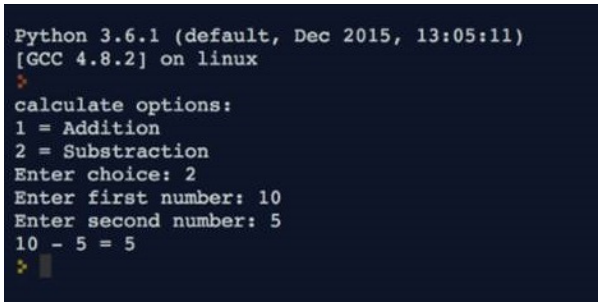



**Fig.5.** Screenshot of the subject´s final recall phase (programming of the calculator).

The test session showed that the study design and procedure of the experiment worked well. Furthermore, the subject gave us some constructive feedback which will be considered in the main experiment.

## References

1. Zimmer, H. D. (2001). Memory for action: A distinct form of episodic memory? Oxford, England: Oxford University Press.
2. Macedonia, M., & Von Kriegstein, K. (2012). Gestures enhance foreign language learning. Biolinguistics, 6, 393–416.
3. Fodor, J. A. (1976). The language of thought. Hassocks, England: Harvester Press.
4. Pylyshyn, Z. W. (1984). Computation and cognition: Toward a foundation for cognitive science. Cambridge, MA: MIT Press.
5. Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
6. Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. Quarterly Journal of Experimental Psychology, 61, 825–850. doi:10.1080/17470210701623605

7. Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. Cognitive Neuropsychology, 22, 455–479. doi:714592738[pii]10.1080/02643290442000310
8. Taylor, L. J., Lev-Ari, S., & Zwaan, R. A. (2008). Inferences about action engage action systems. Brain and Language, 107, 62–67. doi:10.1016/j.bandl.2007.08.004
9. Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. Mind, Brain, and Education, 8(2), 74-88.
10. Pulvermüller, F. (2005). Brain mechanisms linking language and action. Nature Reviews Neuroscience, 6, 576–582.
11. Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. Mind, Brain, and Education, 5(4), 196-211.
12. Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. Human brain mapping, 32(6), 982-998.

# A Cloud-Based Lab Management and Analytics Software for Triangulated Human-Centered Research

Pierre-Majorique Léger*, Francois Courtemanche, Marc Fredette and Sylvain Sénécal

Tech3Lab, HEC Montréal, Montréal, Canada

* pml@hec.ca

**Abstract.** This paper introduces a cloud-based lab management and analytics software platform that we have developed for triangulated human-centred research. The solution is designed to support three main requirements : 1) It enables accurate triangulation of enriched UX measures; 2) It produces triangulated enriched measures in a timely manner; and 3) it helps to generate meaningful recommendations. The application supports the key activities that are required to conduct an enriched UX research project: 1) Designing the UX test; 2) Planning and scheduling the test; 3) Executing the test; 4) Post-processing and triangulating of the collected data; 5) Analyzing and visualizing of the data and; 6) Lab Maintenance. At the time of publication, the application is currently at technology readiness level (TRL) 6 and we are currently conducting beta testing in academic and commercial lab to demonstrate the technology an operational environment.

**Keywords:** UX, neurophysiology, triangulation, lab management system (LMS), heatmaps.

## 1 Introduction

Computerized interfaces are omnipresent in our daily lives and new technologies will continue to transform our lives in the future. What users experience when interacting with an interface is a key determinant of their intention to use and/or reuse a product or service. Research shows that user experience (UX) plays a crucial role in the success or failure of digital products and services [1]. UX is defined as the perceptions and responses that a person has either as a result of using or as a result of anticipating using an information technology (IT) product or service [2]. UX refers to far more than just the ergonomic aspect of an interface, for it encompasses the entire experience as it unfolds before, during, and after the interaction with the interface [3]. The experience is influenced by the usage context as well as by the user's cognitive and affective states.

Industry players and academic researchers are currently facing important scientific and industrial challenges that require urgent attention. Technology evolves rapidly,

and so too must the methods used to evaluate UX. To capture the user's states in authentic usage contexts accurately, research is needed to develop enriched UX measures that enhance existing explicit measures with implicit ones. For example, self-reported measures such as questionnaires could be enhanced by taking account of automatic or subconscious reactions, such as eye movements and psychophysiological reactions, to increase the temporal resolution of measurements. This would capture the full extent of affective states like emotional valence, cognitive states like cognitive workload, and attentional states like divided attention throughout the experience of users. To use these tools and methods effectively, it is necessary to increase the speed and accuracy with which these enriched UX measures can be synchronized, processed, and visualized and to adapt these measures and processes so that they can be used in natural and real-time contexts that are ecologically valid."

The objective of this paper is to introduce a cloud-based lab management and analytics software platform that we have designed to specifically address these research needs. The platform is specifically designed to enable multimodal human computer interaction (HCI) as well as NeuroIS studies involving authentic research stimuli such as that related to the usability of mobile applications. In what follows, we begin by explaining why this tool is needed. After that, we detail the functionalities of this new application, and we conclude by discussing on its current technology readiness level (TRL).

## 2 The requirement: The need to accelerate insight generation from triangulated and enriched UX measures

### 2.1 Enabling triangulation of enriched UX measures

Most of the literature on UX (either in information system (IS) or HCI research) report using explicit measures to assess the user experience [4, 5]. While informative, such measures have drawbacks that could be mitigated by the adjunction of implicit measures. For instance, users can hardly report on their cognitive and emotional states during the interaction without being inherently distracted by the interaction [6] The quasi-exclusive use of explicit measures and data in the field of UX research presents major limitations and potential biases, such as the difficulty of separating the emotional responses to adjacent stimuli, as well as retrospective and social desirability biases [7, 8]. This limitation represents a major industrial problem for diagnosing UX-related issues. Indeed, self-assessed questionnaires, intrinsically, cannot provide, on their own, unbiased longitudinal measures of the UX-related automatic cognitive and emotional mental states during the interaction without the subject being aware of them [4, 9]. For example, questionnaires cannot provide a temporally precise account of the frustration felt by a user when a problem arises at checkout, nor can they distinguish it from the overwhelmingly positive previous impressions left by an online shopping interface at the beginning of their shopping experience.

Research is needed to enrich measures by using implicit measures from which cognitive and emotional constructs can be inferred [4, 10]. Implicit approaches to the investigation of UX can complement current approaches because they address some of the limitations of the explicit measures. First, they enable real-time observation of the user's reactions to technology, as the person interacts with the interface. Second, they permit the capture of subconscious and automatic processes occurring without the individual being aware of them or realizing that they are indeed occurring, thus offering a more complete representation of what actually takes place within the brain, and at what moment. Third, since implicit antecedents need to be captured using methods (e.g., neurophysiological tools) that are different from those used to measure explicit ones (e.g., self-reports), the triangulation of structurally different methods reduces the common-method bias often suspected in UX research [7].

However, triangulation of implicit measures such as multimodal neurophysiological data is usually complex and thus prone to an array of statistical problems, such as high measurement errors and complex dependency patterns among the observations. To our knowledge, there are not commercially available technologies that provide accurate triangulation of multimodal data acquisition. Therefore, we pose our first requirement:

*Requirement 1: The solution needs to enable accurate triangulation of enriched UX measures.*

### 2.2 Processing speed of triangulated enriched UX measures

Now a mainstream approach, Agile Software Development is an umbrella term that refers to a set of methods in which requirements and solutions evolve rapidly over successive iterations. The main objective of this development method is to build useful software less prone to defects and with a shorter time-to-delivery. However, agile methods are not necessarily better at developing usable software that offers a rich experience to potential users [11]. User-centred design can ensure that UX is the focus in software development [12]. But in practice, applying agile development while maximizing UX can present many significant industrial challenges, such as constraints linked to the fast-paced iterations of the software development process [13, 14]. Researchers report on the difficulty that UX professionals have in informing developers in a timely manner, especially when using enriched measures, and show that UX methods are often called upon too late in the development process [15]. Thus, research on process innovation (e.g., innovation at the information system level to fully integrate the research process, end-to-end) is essential to accelerate the pace at which UX research is performed. An enriched UX research cycle that runs over a few weeks or months is simply not satisfactory when software development is scheduled in sprints of 30 days!

The need to accelerate processing speed is also very important in academic research. The imperative to shorten publication cycle time is increasing in the human-centred field such as HCI and psychology. Researchers need tools that will allow them to bring their research discovery sooner in the public space. Also, researchers also benefit from the increase automation in data processing. Relying on manual processing is prone to error and fast research assistant turnover contribute to slowing productive research time.

We therefore pose the second requirement.

*Requirement 2 : The solution needs to produce triangulated enriched measures in a timely manner.*

### 2.2 3 Enabling insight to action

In a usability test, the most crucial take away for a UX designer is not results of the analysis that will come out of the project. The most crucial information is the insight (the aha moment) that one will gain out of this analysis that will to improve the design of an interface. Therefore, the solution must enable the generation of those insights and facilitate the identification of action that can be taken to make profit of those insights. In other words, the solution must inform the design in a meaningful and decisional way and should help, via data visualization methods, to reduce the time required to provide meaningful recommendations to UX professionals.

Here is an example of a novel triangulation and machine learning-based visualization technique that our research team has developed to represent the emotional state of users in a way to support the decision-making process of usability experts. Traditional gaze heatmaps are used in eye tracking as intuitive representations of aggregated gaze data [16]. Their main use is to help answer the question: "Where in the interface do people tend to look?" [17]. In our visualization method, the users' gaze now serves as a means of mapping physiological signals onto the user interface. The resulting heatmaps represent the physiological signals' distribution over the interface, and can help answer the following question: "Where in the interface do people tend to emotionally react?" In a recent publication, we detailed the four steps involved in the creation of physiological heatmaps: inference, normalization, accumulation, and colorization [18, 19]. Using the technique, developers can make a more comprehensive diagnosis of the ergonomic characteristics of an interface and propose an improvement thereof to the design team. This triangulated approach makes it possible to visually analyze users' various emotional states for specific areas of a given interface (e.g., cognitive load combined with emotional valence) [20]. The technique can be used to represent the emotional response to a naturally occurring stimulus on the interface, or to compare the response across multiple versions of the same interface (e.g., A/B testing) [21].

*Requirement 3 : The solution must help to generate meaningful recommendations*

## 4. The solution

We have conducted more than 150 experiments and usability tests over the past 3 years with more than 3000 human subjects. Over time, we have designed processes and systems to meet the three requirements describe in the previous section. The solution we proposed, to meet the requirements described above, a cloud-based lab management and analytics software platform for triangulated human-centred research.

Our solution is a web application developed in a modern programming language designed by Facebook. The application is fully compatible with Noldus Observer XT and Noldus Syncbox. It currently natively handles scientific data acquisition equipment from equipment providers such as Brainvision, Tobii, SMI, Noldus Facereader, and Biopac; several other manufacturers are currently being made compatible. Our solution runs on SAP Hana 2 in-memory analytics which allows for high speed analysis of high dimensional data such as neurophysiological, physiological, and eye tracking data.

There are six key activities that are required to conduct an enriched UX research project: 1) Designing the UX test, 2) Planning and scheduling the test, 3) Executing the test, 4) Post-processing and triangulation of the collected data, 5) Analyzing and visualization of the data. It was designed to support these activities in an integrated manner to increase the processing speed and accuracy of the research data.

The following table details the main functionalities enabled by our solution for each of these activities:

**Table 1.** Main functionalities of the solution

| Activities | Font size and style |
|---|---|
| Designing experiment | Define conditions, events, and stimuli; Define experimental metadata (e.g. file markers naming convention); Define recording equipment; Visualize experimental design. |
| Scheduling experiment | Schedule experimental room; manage reservations; import data from external subject panel system (i.e., Sona System) |
| Executing experiment | Track participant; Write lab notes |

| | |
|---|---|
| Data postprocessing | Upload data from various data collection instruments; Upload data from Observer XT; Validate file according to experimental metadata; Parse data based on event markers; Built-in signal processing with R and python scripts (e.g., HR, HRV, GSR); Automatic data stream synchronization (using sync markers from syncbox). |
| Data analysis | Generate attentional, emotional, and cognitive heatmaps; Generate global experience map; Enable longitudinal and multiple project analysis; Export using Odata to statistical packages from one or multiple projects (e.g, cross-project analyses, compatibilites with third party visualization software such as Tableau software) |

## 5. Discussion and conclusion

At the time of publication, our application is currently in technology readiness level (TRL) 6, i.e., the technology has been tested in a high-fidelity laboratory environment to demonstrates its readiness. In this stage, we have been able to execute more than a dozen usability studies with several neurophysiological signals (including eyetracking, physiological measurements such as electrodermal activities and automatic facial analysis) and to produce actionable results in less than one week (from the moment the first user was tested to the moment where insights were provided to the UX team).

Our next step is to move to TRL 7, which involves demonstration of an actual system prototype in an operational environment. We have several of those tests planned in the next few months to ascertain the capabilities of our solution.

The technology roadmap of our solution is to continue to automate and reduce human intervention in the processing of triangulated and enriched UX measures. Building on new advances in data sciences and artificial intelligence, our goal is to achieve quasi-real-time analysis, enabling to produce results on the same day as a data collection.

## References

1. Kuusinen, K. *Improving UX work in scrum development: A three-year follow-up study in a company*. in *International Conference on Human-Centred Software Engineering*. 2014. Springer.
2. Hartson, R. and P.S. Pyla, *The UX Book: Process and guidelines for ensuring a quality user experience*. 2012: Elsevier.
3. Bevan, N., et al. *New ISO standards for usability, usability reports and usability measures*. in *International Conference on Human-Computer Interaction*. 2016. Springer.

4. de Guinea, A.O., R. Titah, and P.-M. Léger, *Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation.* Journal of Management Information Systems, 2014. **30**(4): p. 179-210.
5. Dimoka, A., P.A. Pavlou, and F.D. Davis, *Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research.* Information Systems Research, 2011. **22**(4): p. 687-702.
6. Ortiz de Guinea, A. and J. Webster, *An Investigation of Information Systems Use Patterns: Technological Events as Triggers, the Effect of Time, and Consequences for Performance.* Mis Quarterly, 2013. **37**(4).
7. De Guinea, A.O., R. Titah, and P.-M. Léger, *Measure for Measure: A two study multi-trait multi-method investigation of construct validity in IS research.* Computers in Human Behavior, 2013. **29**(3): p. 833-844.
8. Malhotra, N.K., S.S. Kim, and A. Patil, *Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research.* Management science, 2006. **52**(12): p. 1865-1883.
9. Tams, S., et al., *NeuroIS-alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research.* Journal of the Association for Information Systems, 2014. **15**(10): p. 723.
10. Riedl, R., F.D. Davis, and A.R. Hevner, *Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement.* Journal of the Association for Information Systems, 2014. **15**(10): p. I.
11. Brhel, M., et al., *Exploring principles of user-centred agile software development: A literature review.* Information and Software Technology, 2015. **61**: p. 163-181.
12. Abras, C., D. Maloney-Krichmar, and J. Preece, *User-centred design.* Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, 2004. **37**(4): p. 445-456.
13. Larusdottir, M., J. Gulliksen, and Å. Cajander, *A license to kill–Improving UCSD in Agile development.* Journal of Systems and Software, 2017. **123**: p. 214-222.
14. Øvad, T. and L.B. Larsen, *How to reduce the UX bottleneck–train your software developers.* Behaviour & Information Technology, 2016. **35**(12): p. 1080-1090.
15. Kuusinen, K., *BoB: A Framework for Organizing Within-Iteration UX Work in Agile Development*, in *Integrating User-Centred Design in Agile Development*. 2016, Springer. p. 205-224.
16. Nielsen, J. and K. Pernice, *Eyetracking web usability*. 2010: New Riders.
17. Wooding, D.S. *Fixation maps: quantifying eye-movement traces*. in *Proceedings of the 2002 symposium on Eye tracking research & applications*. 2002. ACM.
18. Courtemanche, F., et al., *Physiological heatmaps: a tool for visualizing users' emotional reactions.* Multimedia Tools and Applications, 2017: p. 1-28.

19. Georges, V., et al., *Measuring visual complexity using neurophysiological data*, in *Information Systems and Neuroscience*. 2015, Springer. p. 207-212.
20. Courtemanche, F., et al., *Method of and system for processing signals sensed from a user*. 2018, Google Patents.
21. Georges, V., et al. *The Adoption of Physiological Measures as an Evaluation Tool in UX*. in *International Conference on HCI in Business, Government, and Organizations*. 2017. Springer.

# *brownieR*: The R-Package for Neuro Information Systems Research

Sven Michalczyk[1], Dominik Jung[1,2], Mario Nadj[1], Michael T. Knierim[1], and Raphael Rissler[1]

[1]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[2]Karlsruhe Decision and Design Lab (KD2Lab), Karlsruhe, Germany
{d.jung, mario.nadj, michael.knierim, raphael.rissler}@kit.edu

**Abstract.** Neuro-Information-Systems (NeuroIS) research has become an established approach in the information systems (IS) discipline for investigating and understanding user behavior. Our outlined package with the name *brownieR* is a freely-available open source R-package for analyzing NeuroIS data (i.e. the combination of physiological and behavioral data). The central purpose of this work is to instruct researchers how *brownieR* can be used in IS research by providing a practical guide on how to conduct the analysis of bio-physiological data combined with behavioral data (e.g., from the web, experimental tasks, or log files). Further, the article provides an analysis framework and covers the different stages involved in analyzing physiological data.

**Keywords:** NeuroIS · bio-physiology · behavior analytics · web analytics

## 1 Introduction

NeuroIS research has received much attention in recent years due to its potentials to investigate the neural and bio-physiological foundations of cognitive processes and corresponding behaviors, which offer important benefits in understanding the design, development, use, impact, and acceptance of IS [1]. Hereby, NeuroIS is an "interdisciplinary field of research that relies on knowledge from disciplines related to neurobiology and behavior, as well as knowledge from engineering disciplines" [2].
This contributes to a better understanding of the design, but also targets the positive bio-physiological outcomes of IS. The integration of bio-physiological data has many possible use cases in IS, for instance neuro-adaptive IS for emotion regulation [3], decision-support in financial decision-making [4], or emotion management in small, cooperative groups [5]. The NeuroIS methodology has also been proposed to help understand phenomenona like flow [6], or decision inertia [7].
As a foundation, NeuroIS research requires profound knowledge and tools to analyze bio-physiological data. However, only a very limited number of software packages are freely available for the analysis of bio-physiological data [8], especially at the interface of physiological and behavioral data. Having reviewed the R-archive net-

work (CRAN) for the terms ("physiology", "heart rate", "electrodermal activity", "skin conductance", "electroencephalography", "electromyography", "pulse", "blood pressure", "BVP", "PPG", "ECG", "HR", "EDA", "SCR", "SCL", "EMG", "EEG"), only three related packages were found ("biosignalEMG", "eegkit", "RHRV"). A subsequent Google search for the same keys returned two more R-packages listed on Github ("EEK", "EEGUtils"). However, none of them combines IS user behavior data with the analysis of physiological data. Consequently, there remains a need for packages to integratively support NeuroIS research processes, and to provide a methodological toolbox in form of an R package. Such a package should provide a methodological guideline for analyses, but also support established research processes.
This article presents such an R-package. The package that we called *brownieR,* provides user-friendly functions for the selection, preprocessing, and analysis of physiological and behavioral data. It also provides functions to produce graphical representations and descriptive statistics of this data. In this article, our overall aim is to illustrate benefits and application of the package, based on a use case of NeuroIS research. This article is organized along the stages of conducting a NeuroIS analysis, providing a practical guide for how each of the steps can be implemented in R. We illustrate our tutorial with a popular use case in IS, the analysis of behavioral web data for usability tests. Finally, we conclude by summarizing our work.

## 2 The Package *brownieR*

To date, the R package is primarily supporting the data format from the NeuroIS experiment platform *Brownie* [9,10]. However, an integration of further formats is planned. The *Brownie* platform is freely accessible at:

`https://im.iism.kit.edu/1093_1100.php,`

and the source files of the *brownieR* package are available at:

`https://github.com/Fiddleman/NeuroIS.`

In general, the package follows the standard data analysis process as proposed by Wickham and Grolemund [11], which is illustrated in figure 1. First of all, the data needs to be imported into R. Before we can start the modeling or visualizing phase, the data has to be transformed. Usually, before a model is built, we explore the data by plotting visualizations and by looking at descriptive statistics. After these steps, findings can be communicated. In this technical article, we follow this general process to structure the analysis of bio-physiological data. Furthermore, we use the user-called-functions for most of the naming of common functions in R like *plot()*, or *summary()*. Our analysis starts by using the *import()* function, which adds all log-files in a specified folder as objects to the global environment.
If an analysis on physiological data has to be conducted, a marker object for the relevant events has to be created. For that purpose we rely on the *marker()* function. To

get an aggregated view on the data, the *summary()* function can be called on each of the objects. This view provides insights for plotting the data. In an iterative manner, we can dive deeper into the data by running the *summary()* and *plot*() function again. If we conducted the experiment on a website, we provide the *as_clickstream()* function to transform the data for modelling as Markov Chain, using the *clickstream* R-package by Scholz [12].

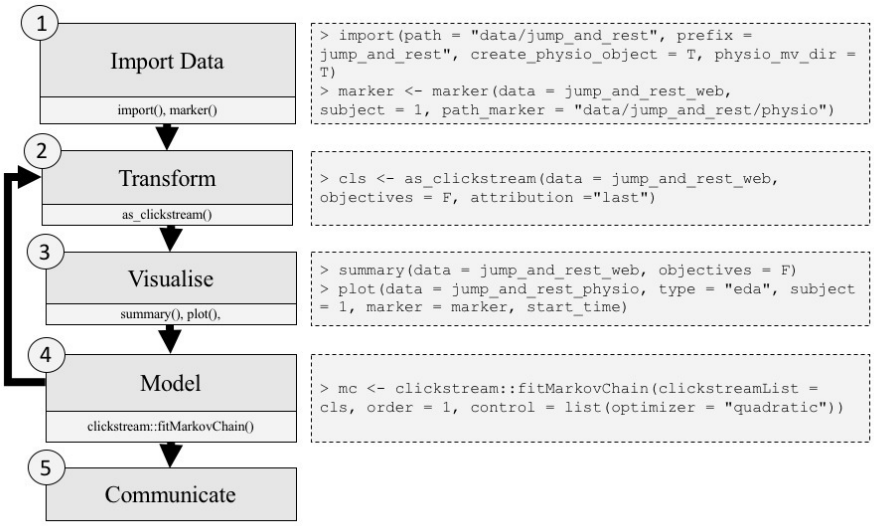



**Fig. 1.** The NeuroIS analysis process of this R-package based on Wickham [11]

## 3 Use Case: Analysis of Behavioral Web Data for Usability

In this chapter, we illustrate the package in action by relying on a popular use case in IS, the analysis of behavioral web data for usability tests. The data for our example is generated by the following setting: A user clicks randomly on different elements at the website http://iism.kit.edu. In order to introduce variance to the physiological data the user started jumping up and down for approximately 10 seconds from time to time. ECG, BVP and EDA data as well as web log-files are tracked by *Brownie* during the session. On this basis, several common research questions in NeuroIS research can be derived, which serve as a "route map" in the upcoming subchapters to explain our implementation:

1. Do central pages exist in terms of duration or click frequency?
2. At which pages did the user jump?
3. Which elements on a central page did the mouse last on?
4. What is the probability to go from one page to another, based on the user?

### 3.1 Data Import and Transformation

Depending on the kind of experiment, *Brownie* tracks the data in three different structures: *web* (web data), *physio* (bio-physiological data) and *exp* (experimental data). *Physio* data can be collected in every experiment in case that sensors are attached to the participant. Additionally, only *web* or *exp* data can be tracked depending on whether the experiment runs in the browser or is implemented in Java using *Brownie*. For our example implementation, we tracked solely *web* and *physio* data.
Firstly, we use the *import()* function to read the data into R from a target folder with the code outlined below:

```
library("brownieR")
import(path = "data/jump_and_rest", prefix =
"jump_and_rest",
      create_physio_object = T, physio_mv_dir = T)
```

The *import()* function has several arguments that are important for further analysis. We only have full capabilities if we set the arguments *create_physio_object* and *physio_mv_dir* to true. The drawback is that a *physio* object can be very large. For example, ten minutes of physiological data for ten users are approximately 500 Megabyte large (given the 1000Hz sampling rate here). The name of the objects can be specified with the argument prefix. The *import()* function does not only import data into R. It also detects the data-type for each file and combines files from different users to the related object class. Additionally, the function removes empty columns to reduce size, whereas all time columns are converted to the common data format *POSIXct*.

### 3.2 Visualization and Data Understanding

After importing the data, we have a list of class type *physio* and another one of class type *web* in our global environment. To answer our first question about frequently clicked pages, we run the function *summary()* on the web object.

```
summary(data = jump_and_rest_web, objectives = F)
```

The argument *objectives* can be used to calculate web conversion rates. The appropriate use can be reviewed in the built-in R documentation by calling the function *?summary.web*.

**Table 1.** Output of the *summary()* function (for our example data).

| | URL | Impression | Session | Duration | Imp_per_Session |
|---|---|---|---|---|---|
| 1 | kd2lab.kit.edu/21.php | 2 | 1 | 10.6 | 2 |
| 2 | kd2lab.kit.edu/28.php | 2 | 1 | 10.0 | 2 |
| 3 | kd2lab.kit.edu/ | 1 | 1 | 5.7 | 1 |
| 4 | kd2lab.kit.edu/index.php | 1 | 1 | 9.3 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 5 | kd2lab.kit.edu/26.php | 1 | 1 | 7.0 | 1 |

The column *Impression* shows, how often a URL (i.e. a single page), is shown to participants during the experiment. The column *Session* provides information on how many users invoke a URL. The column *Duration* offers the average length of stay at a URL over all impressions in seconds. In our example, the page 21.php has the longest durations of 10.6 seconds and with 28.php the most impressions. These two pages can be seen as central.
The *summary()* function can be run on a *physio* object as well. As we overload the functions with R's S3 methods, the function has a completely different output than the summary from base R. First of all, we need to create a marker object, which is necessary to see the URL changes (or screen changes in case of *exp* data) along the physio data. The command refers to the following:

```
marker <- marker(data = jump_and_rest_web,
subject = 1, path_marker = "data/jump_and_rest/physio")
```

We need to specify for which user the marker will be created with a subject. Every user has its unique marker file as all users execute screen or URL changes at different times. Additionally, we can specify a folder with *path_marker* where a marker file will be created. The creation of a marker file is only necessary in case you want to plot ECG data. Doing so, the marker file needs to be in the same folder as the *physio* data, usually a sub-folder named 'physio'.
Now we can run the summary on the *physio* data. Therefore, we need to specify for which type of physio data, ECG, EDA or BVP shall be summarized. Currently, *summary()* is only available for EDA data.

```
start_time <-
min(jump_and_rest_web[jump_and_rest_web$SUBJECT_ID_SUBJEC
T == 1]$Time, na.rm = T)
summary(jump_and_rest_physio, type = "eda", subject = 1,
marker = marker, start_time = start_time)
```

Before running the summary, we need to determine the *start_time* of the experiment for the user. The *start_time* usually refers to the minimal time in *web* or *physio* data for the selected user (subject).
The output of the function is the following:

**Table 2.** Structure of the current physio data, after pre-processing.

| | Time | Event | Average EDA |
|---|---|---|---|
| 1 | 0,76 | https://www.kd2lab.kit.edu/ | 89.66 |
| 2 | 5,74 | https://www.kd2lab.kit.edu/index.php | 89.94 |
| 3 | 14,75 | https://www.kd2lab.kit.edu/21.php | 76.34 |

| | | | |
|---|---|---|---|
| 4 | 25,74 | https://www.kd2lab.kit.edu/28.php | 87.60 |
| 5 | 35,76 | https://www.kd2lab.kit.edu/21.php | 71.36 |
| 6 | 45,74 | https://www.kd2lab.kit.edu/26.php | 86.31 |
| 7 | 52,75 | https://www.kd2lab.kit.edu/28.php | 86.99 |

With this view, we can answer the second question at which specific pages the user might have been jumping. It seems noticeable that the average EDA is clearly smaller at the page 21.php for the time of 14.75 and 35.76.
Furthermore, the package provides another visualization of this finding. For that purpose, we run the *plot()* function on the *physio* data for this user.

```
plot(data = jump_and_rest_physio, type = "eda",
subject = 1, marker = marker, start_time)
```

Alongside the *summary()* function, *type, subject, marker* and *start_time* need to be provided as arguments. We again plot the EDA data. Furthermore, plotting ECG data is possible by changing some arguments:

```
plot(jump_and_rest_physio, type = "ecg", subject = 1,
marker = "data/jump_and_rest/physio/marker1.csv",
start_time, physio_unisens_dir =
"data/jump_and_rest/physio/")
```

In the plot for EDA data, we get a visualization of what we have already investigated by running the *summary()* function.

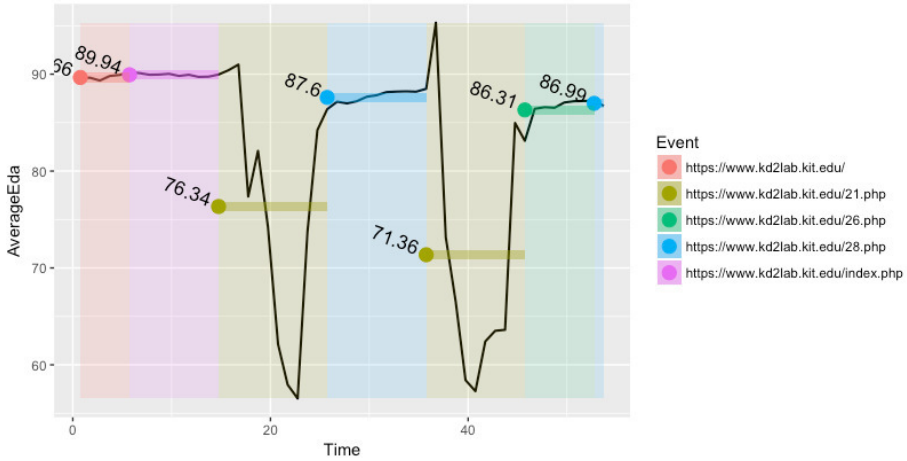



**Fig. 2.** EDA plot visualization of the *brownieR* package

### 3.3 Modeling and Communication

In this chapter, we focus on the third question: Which elements on a central page did the mouse last on longest? We can answer this question by modeling the user behavior as a Markov Chain and by running the *plot()* function on the resulting web object.

```
web_summary <- summary(data = jump_and_rest_web, objec-
tives = F)
urls <- web_summary\$URL
take_screenshot(urls = urls)
plot(jump_and_rest_web, url = urls[3], type = "motion",
subject = 1,
alpha = 0.1, size = 3, color = "purple")
```

Beforehand, some preparations are required: Firstly, we assign the output of the *summary.web()* function to an object to get a list of all URLs in the experiment. Afterwards, we make screenshots with the function *take_screenshots*. This function takes screenshots of all provided URLs and saves them in a sub-folder 'screenshots' in the working directory. Then, the *plot.web()* function has other arguments than the *plot.physio()* function. The arguments *alpha*, *size* and *color* change the appearance of the plot. As every time only one URL can be plotted, we need to specify the URL in the argument *url*. This is different with the argument *subject*. If we have an experiment with more than one user (which is usually the case), we can specify several or all users, by relying on a vector containing the user numbers. This is particularly useful to observe an area of interest, that all users moved their mouse over. In our example, the mouse was moved from the navigation bar to the lab address, which is an indicator that the user was interested in the location of the lab. Further, we can plot clicks on a page as well.

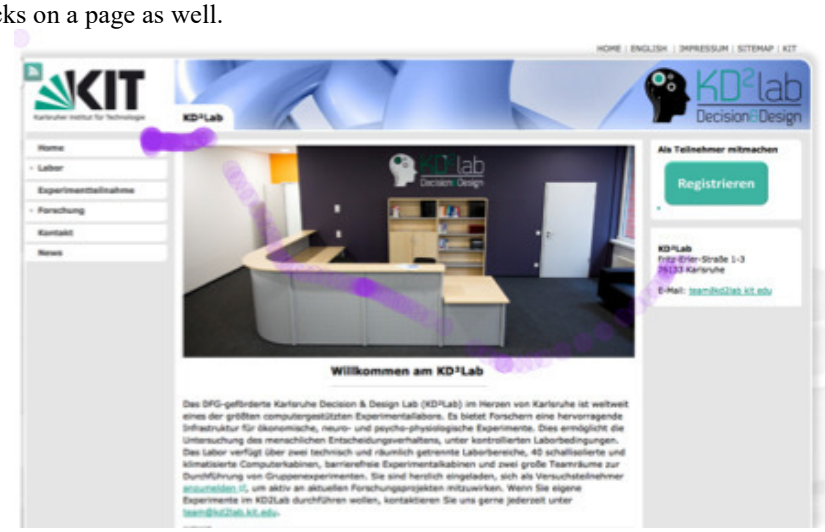



**Fig. 3.** Mouse movement visualization of the NeuroIS R package

To answer question four (transition probabilites from one page to another), we propose the usage of the package *clickstream* by Scholz [12].

```
cls <- as_clickstream(data = jump_and_rest_web,
objectives = F, attribution ="last")
mc <- fitMarkovChain(clickstreamList = cls, order = 1,
control = list(optimizer = "quadratic"))
summary(mc)
plot(mc, order = 1)
```

First of all, we need to transform the web data into a proper format for the *clickstream* package. For this purpose, the package provides a function called *as_clickstream()*. We have the option to specify goal pages of the website. A goal page of a website could refer to the thank-you-page after a successful purchase. As we do not offer a final end page in our example, we set the argument objectives to *false*. Additionally, we can define an attribution model with the argument attribution (first or last are possible values). By applying the function *fitMarkovChain()* to the output of the *as_clickstream()* function, we start to use the clickstream package and build a first order markov chain by using a quadratic optimizer. The model is saved to the class *mc*. On this basis, we can apply a summary or a plot function on this model to visualize it. For instance, the plot of the markov model allows visualizing the transition probability from one page to another.

## 4 Conclusion

With the advancements of NeuroIS research to examine and understand cognitive and physiological processes and associated user behaviors, this interdisciplinary research domain have become increasingly relevant at the nexus of IS artefact design in both academic and business communities. However, designing such artefacts and understanding the underlying phenomena, requires a profound knowledge and tool support to analyze the corresponding bio-physiological data, especially at the interface of behavioral data. By creating an R-package for the analysis of physiological and behavioral data, this piece of research provides a foundational step towards a toolbox supporting the need for an overarching NeuroIS research process. However some limitations are worth noting, in particularly the integration of more complex datatypes is still missing, and the toolbox does not support more complex data like fMRI data or eyetracking data at the moment. We are looking forward to integrate that in a future version of our toolbox.

In particular, we hope that this freely-available open source package can serve as a methodological guideline and analysis framework for scholars and practitioners in the broader landscape of NeuroIS research covering the various stages involved in analyzing bio-physiological data in combination with user behavior data.

## References


[1] Riedl R, Léger PM (2016). Fundamentals of NeuroIS. Springer.

[2] Riedl R, Banker RD, Benbasat I, Davis FD, Dennis AR, Dimoka A, Gefen D, Gupta A, Ischebeck A, Kenning P, et al. (2010). "On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009." CAIS, 27, 15.

[3] Astor PJ, Adam MT, Jerčić P, Schaaff K, Weinhardt C (2013). "Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation." Journal of Management Information Systems, 30(3), 247–278.

[4] Jercic P, Astor PJ, Adam MT, Hilborn O, Schaaff K, Lindley C, Sennersten C, Eriksson J (2012). "A serious game using physiological interfaces for emotion regulation training in the context of financial decision-making." In Proceedings of the 20th European Conference on Information Systems (ECIS).

[5] Knierim MT, Jung D, Dorner V, Weinhardt C (2017a). "Designing Live Biofeedback for Groups to Support Emotion Management in Digital Collaboration." In International Conference on Design Science Research in Information Systems, pp. 479–484. Springer.

[6] Knierim MT, Rissler R, Dorner V, Maedche A, Weinhardt C (2017b). "The psychophysiology of flow: A systematic review of peripheral nervous system features." In Gmunden Retreat on NeuroIS.

[7] Jung D, Dorner V (2017). "Decision Inertia and Arousal: Using NeuroIS to Analyze Bio-Physiological Correlates of Decision Inertia in a Dual Choice Paradigm." In Gmunden Retreat on NeuroIS.

[8] Rosenberg, D. M., & Horn, C. C. (2016). Neurophysiological analytics for all! Free open-source software tools for documenting, analyzing, visualizing, and sharing using electronic notebooks. Journal of neurophysiology, 116(2), 252-262.

[9] Hariharan, A., Adam, M. T. P., Lux, E., Pfeiffer, J., Dorner, V., Müller, M. B., & Weinhardt, C. (2017). Brownie: A platform for conducting NeuroIS experiments. Journal of the Association for Information Systems, 18(4), 264.

[10] Jung, D., Adam, M., Dorner, V., & Hariharan, A. (2017). A practical guide for human lab experiments in information systems research: A tutorial with Brownie. Journal of Systems and Information Technology, 19(3/4), 228-256.

[11] Wickham H, Grolemund G (2016). "R for data science."

[12] Scholz M (2014). "Analyzes clickstreams based on Markov chains." R package version, 1(2).

# Enhancing the Implicit Association Test: A Four-Step Model to Find Appropriate Stimuli

Gerhard Brenner[1], Monika Koller[2], Peter Walla[3,4,5]

[1] University of Applied Sciences, Johannes Gutenberg-Straße 3, 2700 Wiener Neustadt, Austria
gerhard.brenner@fhwn.ac.at
[2] WU Vienna, Welthandelsplatz 1, 1020 Vienna, Austria
monika.koller@wu.ac.at
[3] CanBeLab, Department of Psychology, Webster Vienna Private University, Praterstrasse 23, 1020 Vienna, Austria
peter.walla@webster.ac.at
[4] School of Psychology, Newcastle University, Newcastle, Callaghan, Australia
[5] Faculty of Psychology, University of Vienna, Vienna, Austria

**Abstract.** The Implicit Association Test (IAT) is a promising tool to assess implicit attitudes. Next to neuroscientific methods applied within the field of NeuroIS the IAT helps to overcome limits of traditional approaches, such as self-report studies. Introduced 20 years ago, it has been developed further within subsequent years. However, hardly any attention has been paid to optimize the stimuli sets. This is unfortunate, as if the time span participants need to decode the stimuli varies across the IAT, or if the subjects do not understand the stimuli equally, reaction times can be biased. As an IAT includes 120 measuring points per subject such biases might potentiate across all participants. The results might be biased and neither the researchers nor the participants would recognize such confounding effects. Thus, we focus on the time span between stimulus onset and response and develop a four-step model to create an optimized stimuli set including (1) brainstorming, (2) forming & performing (i.e. pretesting), (3) backward-brainstorming and (4) informing & interviewing.

**Keywords:** Implicit Association Test, Implicit Measures, Indirect Measures.

## 1 Introduction

In both, academic as well as applied information systems research, measuring attitudes and emotions, e.g., associated with system use, has been important for decades. However, especially when it comes to the measurement of attitudes at an implicit level, traditional approaches have their limits. The field of NeuroIS has offered promising ways to help in this regard. For example, fMRI applications for instance to investigate psychological phenomena such as Dimoka (2010) did on trust [1] (also see a review by Riedl and Javor on this topic [2]) or other physiological approaches, such as Walla and Koller (2015) did on using startle reflex modulation [3] have been introduced to measure attitudes at an implicit level in the context of information systems.

Besides methods from neursocience and/or psycho-physiological methods, the Implicit Asssociation Test (IAT) has proven to be an alternative instrument to capture implicit attitudes in multiple disciplines. Today, the IAT is one of the most frequently applied methods to assess implicit mental processes such as implicit attitudes [4], not only in the field of social psychology, but also in consumer research [5,6,7]. Particularly, implicit attitude measurement methods are applied to overcome unwanted effects when participants answer questions of questionnaires. Such effects are often caused by social desirable answering, impression management, self-deviation, the incapability of introspection, and finally by low memory capacity [7]. Pretty often, when both implicit and explicit attitudes are measured, the results differ significantly from each other [9,10]. Hence, in information systems research, applying the IAT could help to gain further insights whenever implicit attitudes are measured. For instance, when it comes to investigate current phenomena like the acceptance of new technologies, technostress or other negative consequences of overconsumption of digital technologies.

During the years following the introduction of the IAT as a method to measure implicit attitudes, the IAT has continuously been developed further [11,12,13,14,15,16]. Greenwald, Banaji, and Nosek [10] suggested an "improved scoring algorithm" as well as an enhanced IAT procedure. The new algorithm included the computation of a D-measure putting the mean reaction times in relation to standard deviations. The new IAT procedure comprised seven task sequences ("blocks"), compared to only five blocks in the original version of the IAT [17].

Compared to the number of blocks featured as well as the improved scoring algorithm, the IAT stimuli set has gained only little attention in previous research. This is unfortunate, as the stimuli set is crucial for the entire measurement approach incorporated in the IAT. As one example, Fiedler, Messner, and Bluemke [18] pointed out that single stimuli might be misunderstood by IAT subjects. Although, the creation of the stimuli set has already been named as critical for the entire procedure, comprehensive suggestions for improvement of the procedure are still missing.

## 2 The Implicit Association Test

In an IAT, the subjects are confronted with stimuli, either words or pictures representing four categories: Two of them are *target categories* – these are categories that are the targets to be measured (e. g. black vs. white people to assess prejudices of white over black people); and there are two *attribution categories,* representing emotional valence (positive vs. negative). [17]

The principle underlying the IAT is that people quickly respond (e.g., pressing a left or right key) to stimuli representing the *target categories* whenever these categories and the *positive* or *negative* emotional words *(attribution categories)* are implicitly understood congruently. On the other hand, the reaction times will extend when the underlying concepts do not fit. According to Greenwald et al. (2003) the IAT is performed by completing seven blocks (see figure 1). Reaction times are relevant for measurement in blocks 3, 4, 6, and 7. Reaction times following congruent stimuli

(supposed to be shorter) are subtracted from reaction times following incongruent stimuli and subsequently divided by their pooled standard deviation [11].
Stimuli sets used in the IAT are usually created by brainstorming and pretests (e.g., Gattol, Sääksjärvi, and Carbon [19]). However, this procedure might be too simple and produce misunderstandings.

| task | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **categories** | **target categories** (20 trials) | **attribution categories** (20 trials) | **categories combined** (20 trials) | **categories combined** (40 trials) | **target categories (reversed)** (40 trials) | **categories combined (reversed)** (20 trials) | **categories combined (reversed)** (40 trials) |
| instructions | black people --> press "d"<br>white people --> press "k" | positive --> press "d"<br>negative --> press "k" | black people --> press "d"<br>positive --> press "d"<br>white people --> press "k"<br>negative --> press "k" | black people --> press "d"<br>positive --> press "d"<br>white people --> press "k"<br>negative --> press "k" | white people --> press "d"<br>black people --> press "k" | white people --> press "d"<br>positive --> press "d"<br>black people --> press "k"<br>negative --> press "k" | white people --> press "d"<br>positive --> press "d"<br>black people --> press "k"<br>negative --> press "k" |

**Fig. 1.** Seven IAT blocks: Combined task blocks 3, 4, 6, and 7 are relevant for measurement [11].

Since the IAT has been introduced by Greenwald et al. in 1998 [17], researchers also developed an IAT to generate results based on large numbers of participants. Greenwald, Banaji, and Nosek founded the "project implicit" on the Harvard University website (https://implicit.harvard.edu) to initiate web-based IATs with large numbers of participants [20]. In light of the previously mentioned issues of wrongly encoded stimuli [18], we argue that comprehensively preparing all single elements of an IAT is of utmost importance to avoid any biased or misleading results.

## 3 Four-Step Model

To avoid wrongly decoded stimuli we developed a four-step model to find stimuli for the four IAT categories encompassing the steps: (1) brainstorming, (2) forming & performing, (3) backward-brainstorming, and (4) informing & interviewing.

### 3.1 Criteria to Find Stimuli

The criteria the stimuli should meet are determined by their fit with the categories they represent: *First of all,* the stimuli should be *decoded* as fast as possible by the subjects and within the same time span, so that the difference of the reaction time span is determined by its evaluation only. What is ought to be measured by the IAT is the evaluation of a stimulus. However, the time span between stimulus onset and response (pressing a key) encompasses (a) identifying the stimulus, (b) understanding its meaning (including a phase of potentially wrong understanding), and (c) its evaluation. If a stimulus is decoded differently because of certain decoding problems (see as a hypothetical example depicted below), the evaluation of that IAT stimulus is altered (see figure 1).

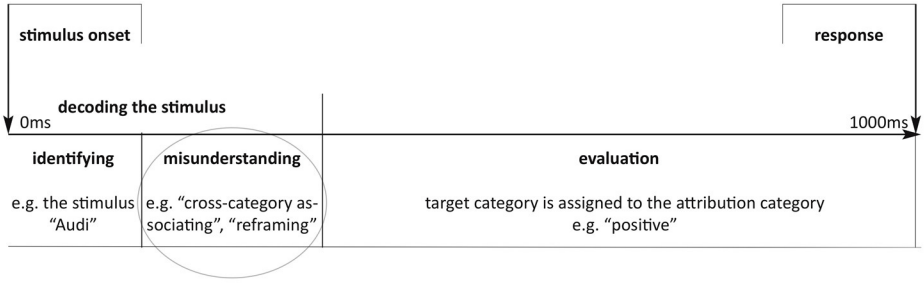



**Fig. 2.** Timeline (e.g., 1000ms) between stimulus onset and response (e.g. press key "d").

*Second, if possible,* the stimuli representing the *target categories* should not be emotionally loaded. For instance, if the target category "GDR" *(German Democratic Republic)* is represented by the word "torture", it is the valence of the stimulus itself, which makes the response fitting the negative valence, but not its evaluation of the target category "GDR" – no matter whether the word "torture" can be representative for this category or not.

*Third,* the stimuli should be as specific as possible for the categories they represent. The word "motorcycle" for instance is not necessarily understood as "non-environmental" as suggested by the IAT of Gattol et al. [19]. Some participants might be bikers themselves, and thus might not associate "motorcycle" per se with "non-environmental" vehicles.

*Fourth,* the researchers should seek to eliminate potential reframing effects. The problem of reframing occurs when – as in Fiedler et al's example – the category "Turkey" in one situation is framed as politically problematic, and in another situation it is framed as a popular holiday destination [18].

*Fifth,* the problem of "cross-category association" [18] occurs, when a stimulus fits both target categories or both attribution categories. For instance, the female prename "Katie", used by Greenwald et al. for *black* people, could be associated with *white* people just as well, at least if the IAT was performed today [17].

*Sixth* and finally, stimuli triggering "like-me" associations might be clearly positive, whereas "unlike-me" associations might be negative. Gattol et al. [19] used "Switzerland" as a stimulus representing the category "safety". However, the word "Switzerland" might be perceived more positively by participants coming from Switzerland, because of "like-me" associations, than for instance Germans or Austrians would, because of their "unlike-me" associations with "Switzerland".

To summarize, the aim of the present paper is to suggest a procedure for generating valid sets of stimuli through avoiding the six potential pitfalls outlined.

### 3.2 Four Steps

We suggest a 4-step model, in which those criteria and potential pitfalls are addressed: In *step 1,* a basic set of IAT stimuli is created by carrying out a *brainstorm-*

*ing* process. However, researchers are advised not to brainstorm on their own. They should include members of their planned sample population. In a first step, we recommend to run the brainstorming in written form by each team member on his or her own. In a second step, the team members should read out loud the words they have found, compare them and then add further associative words.

In *step 2,* the so called "forming and performing" step, problematic associative words are removed from the stimuli sets. As mentioned above, problems particularly occur with *emotionally valenced* stimuli representing one of the target categories. The same is true for *unspecific* stimuli (representing both, target and attribution categories), for stimuli potentially eliciting *reframing effects,* for stimuli potentially eliciting *cross-category associating effects,* and for stimuli which potentially elicit either a "like-me" or an "unlike-me" effect. When this procedure is completed, we recommend a first pretest. After each IAT, the participants are interviewed with the focus being on each of the stimuli [20]. If the stimulus set has not been reduced to a number of stimuli which is feasible for a pretest to be carried out (above 10 stimuli for each category, because too many stimuli might confuse the participants), after the forming procedure, pretesting can be postponed to step 3.

In *step 3,* a couple of new participants (again reflecting the planned sample population) is asked to undergo a "backward-brainstorming" process. The words which have been found in steps 1 and 2 are presented to the subjects and the participants are asked to spontaneously disclose what they associate them with. If they associate the words with the corresponding category, the words work properly – the quicker (first or later announced), the better. As a result, these associations are categorized by the criteria "first announcement" (category being announced as a first association) and "announced later on" (category being announced as a second or further association). We recommend to perform a couple of IAT pretests after having completed the backward-brainstorming procedure, to find out whether misunderstandings still might occur.

*Step 4,* finally, is carried out when the IAT is already designed for the experiment. It encompasses "informing" the participants before starting the IAT and "interviewing" them after having completed it. "Informing" means to tell the subjects which words represent which category and asking them to read those words out loud to make sure that they know the words and that they will associate them with the correct categories. This is kind of a learning procedure. In the interviews, after having completed the IAT, potential misunderstandings can be revealed if they have occurred despite the researchers best efforts to avoid them.

Particularly, the learning procedure ("informing") in step 4 allows the researcher to limit the presentation time of each stimulus to as low as 150ms. This short presentation time and an inter-stimulus interval of 2000ms guarantee affective reactions and should avoid unwanted cognitive processes within the subjects' minds. Cognitive processes might distract the affective nature of the IAT and should therefore be inhibited. Moreover, many IAT researchers give their participants the opportunity to correct false category assignments. If for instance a participant has pressed the left key instead of the right-one, he or she is given time to press the correct key. However, we suggest that the participants should not have the opportunity to correct erroneous reactions. Those errors are treated the same way as missing values are (mean of cor-

rect reaction times plus 600ms), as suggested by Greenwald et al. [11]. By applying the four-step model to find stimuli the number of erroneous reactions can be reduced to a minimum. Error rates are one of the critical aspects of the IAT. Westgate et al. for example excluded data from participants who had produced an overall error rate of 30 per cent or more, or an error rate of individual blocks of 40 per cent or more [21]. We endorse this suggestion.

## 4 Conclusion

For researchers investigating attitudes, the IAT is a promising tool to overcome problems of social desirable answering, impression management, self-deviation, the incapability of introspection, and low memory capacity. However, if the IAT is not prepared diligently and performed carefully, its results might be biased. In this article, we focused on the time span between stimulus onset and response to the stimulus by pressing a key. Our proposed four-step model has the potential to optimize word stimuli sets to keep the decoding time span constant across all participants. If single decoding sequences are contaminated by distracting effects such as cross-category association effects, these biases may cause multiple contaminations of hundreds of milliseconds, which potentiate across a total of 120 measuring points in blocks 3, 4, 6, and 7. The crucial thing is that nobody will realize these contamination effects, because neither the researchers will be aware of what is going on within the subjects' minds nor the participants themselves will be aware of it.

The four-step model we developed is an exploratory suggestion on how the stimuli set used in an IAT can be created in a valid and comprehensive manner. Besides the merits of the suggested approach, there are still limitations subject to further research. For instance, the backward-brainstorming sequence might be calculated in a more sophisticated way by developing an algorithm to consider the weight of first-announcements and later-on-announcements.

### References


1. Dimoka, A.: What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. MIS Quarterly, 34, 2, 373–396 (2010)
2. Riedl, R., Javor, A.: The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. Journal of Neuroscience, Psychology, & Economics, 5, 2, 63–91 (2012)
3. Walla, P., Koller, M.: Emotion is not what you think it is: Startle reflex modulation (SRM) as a measure of affective processing in NeuroIS. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. (eds.) Information Systems and Neuroscience, Gmunden Retreat on Neuro IS 2015, pp. 181–186. Springer International Publishing Switzerland, Cham, Heidelberg, New York, Dortrecht, London (2015)
4. Gawronski, B., De Houwer, J.: Implicit measures in social and personality psychology. In: Reis, H.T., Judd, C.M. (eds.) Handbook of Research Methods in Social and Personality Psychology, pp. 283–310. Cambridge University Press, New York (2014)

5. Gregg, A.P., Klymowsky, J., Owens, D., Perryman, A.: Let their fingers do the talking? Using the Implicit Association Test in market research. International Journal of Market Research, 55, 4, 23–39 (2015)
6. Bosshard, S., Bourke, J.D., Kunaharan, S., Koller, M., Walla, P.: Established liked versus disliked brands: Brain activity, implicit associations and explicit responses. Cogent Psychology, 3, 1 (2016)
7. Walla, P., Koller, M., Brenner, G., and Bosshard, S.. Evaluative conditioning of established brands: implicit measures reveal other effects than explicit measures. Journal of Neuroscience, Psychology and Economics, 10, 1, 24–41 (2017)
8. Fazio, R.H., Olson, M.A.: Implicit measures in social cognition research: Their meaning and use. Annual Review of Psychology, 54, 297–327 (2003)
9. Gawronski, B., Strack, F.: On the propositional nature of cognitive consistency: Dissonance changes explicit but not implicit attitudes. Journal of Experimental Social Psychology, 40, 535–542 (2004)
10. Rydell, R.J., McConnell, A.R., Mackie, D.M., Strain, L.M.: Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. Psychological Science, 17, 11, 954–958 (2006)
11. Greenwald, A.G., Banaji, M.R., & Nosek, B.A.: Understanding and using the Implicit Association Test: An improved scoring algorithm. Journal of Personality and Social Psychology, 85, 197–216 (2003)
12. Greenwald, A.G., Nosek, B.A.: Health of the implicit association test at age 3. Zeitschrift für Experimentelle Psychologie, 48, 85–93 (2001)
13. Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., Banaji, M.R.: Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of Personality and Social Psychology, 97, 1, 17–41 (2009)
14. Lane, K.A., Banaji, M.R., Nosek, B.A., Greenwald, A.G: Understanding and using the Implicit Association Test: IV – What we know (so far) about the method. In: Wittenbrink, B., Schwarz, N. (eds.) From implicit measures of attitudes, pp. 59–102. The Guilford Press, New York (2007)
15. Nosek, B.A., Greenwald, A.G., Banaji, M.R.: The Implicit Association Test: II. Method variables and construct validity. Personality and Social Psychology Bulletin, 31, 166–180 (2005)
16. Nosek, B.A., Greenwald, A.G., Banaji, M.R.: The Implicit Association Test at age 7: A methodological conceptual review. In: Bargh, J.A. (ed.) Automatic processes in social thinking and behavior, pp. 265–292. Psychology Press, New York (2007)
17. Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K.: Measuring individual differences in implicit cognition: The Implicit Association Test. Journal of Personality and Social Psychology, 74, 1464–1480 (1998)
18. Fiedler, K., Messner, C., Bluemke, M.: Unresolved problems with the „I“, the „A“, and the „T“: A logical and psychometric critique of the Implicit Association Test (IAT). European Review of Social Psychology, 17, 74–147 (2006)
19. Gattol, V., Sääksjärvi, M., Carbon, C.-C.: Extending the Implicit Association Test (IAT): Assessing consumer attitudes based on multi-dimensional implicit associations. PloS ONE 6, DOI: 10.1371/journal.pone.0015849 (2011)
20. Project Implicit, https://implicit.harvard.edu, last accessed 2018/03/21.
21. Westgate, E.C., Riskind, R.G., Nosek, B.A.. Implicit preferences for straight people over lesbian women and gay men weakened from 2006 to 2013. Collabra, 1, 1, 1–10 (2015)

# Facebrain: A P300 BCI to Facebook

Ben Warren, Adriane B. Randolph

Kennesaw State University, Department of Information Systems, Kennesaw, Georgia, USA
ben@warr.io, arandol3@kennesaw.edu

**Abstract.** Facebrain is a novel brain-computer interface utilizing the P300 signal as input for interacting with the Facebook social media platform. Electroencephalography along with the open-source BCI2000 software suite is used for both obtaining and processing the signals. Additionally, BCPy2000, an add-on allowing BCI2000 modules to be written in the scripting language Python, is utilized to allow for rapid interface generation, promoting extensibility, and a cross-platform solution. Users are able to select basic Facebook operations via a P300 matrix and then activate a P300 speller as needed for text input. Overall, the purpose of the system is to allow functional, hands-free, and voiceless access to Facebook's main features including, but not limited to, searching for and adding friends, making posts, using the chat system, and browsing profiles.

**Keywords:** Brain computer interface · EEG · P300 · BCI2000 · Python · Facebook · communication.

## 1 Introduction

Facebook (http://facebook.com), the social media platform, is a prime example of the role of technology in aiding communication, yet it is virtually inaccessible to those with severe motor impairments. Facebrain is an application designed to provide access to the major functionality of the popular social media platform using the principles behind P300 stimulus presentation. Such access would allow individuals who are locked-in the ability to proactively interact with friends and family using a medium now common to over a billion users [1]. For example, a survivor of a spinal cord injury resulting in locked-in syndrome who was a previous user of Facebook may retain the ability to communicate with loved ones seamlessly via the same medium, though without the need for mechanical or vocal input into a computer. Those who have had long-term impairments would be given an avenue to stay informed and communicate with others in a way previously unavailable to them.

Brain-computer interfaces have provided a platform for individuals with long-term motor impairments to communicate and control their environments using neural input alone [2, 3]. Individuals with locked-in syndrome are paralyzed and unable to speak and yet otherwise cognitively in-tact and in need of a non-traditional communication

system [4]. These interfaces have been shown to significantly increase quality of life and promote social inclusion [5, 6]. Further, researchers have shown the benefits of web-based and multimedia applications effectively accessed via P300 by healthy participants and participants with locked-in syndrome [7]. Here, we present a P300-based interface that integrates a tool for social inclusion in a manner appropriate for individuals with severe motor disabilities.

P300 is an event-related potential (ERP) present in an electroencephalograph (EEG) that corresponds with attention to infrequent presentation of a stimulus. Researchers have taken advantage of this signal to implement novel systems capable of discerning when a desired stimulus is indicated among an array of alternates. Spelling systems have been one of the most promising applications, allowing individuals to compose words and sentences by simply focusing on a particular character amongst a matrix of flashing letters [3].

We used the classic P300-spelling framework as a basis for our design where others have shown up to 90% accuracy with healthy participants on similar interface paradigms [7]. In addition, a visual interface, such as described in the following, has been shown to provide better results than an auditory paradigm for eliciting a P300 response with severely disabled participants [8]. This work presents an interface similar to most BCIs designed for synchronous use where a caregiver or research starts the system and the participant may then use it during the prescribed period of time [9].

## 2 System Design

### 2.1 BCI2000/BCPy2000

BCI2000 is an open-source application written in C++ used to design experiments for BCI research [10, 11]. BCI2000 supports a modular system consisting of three communicating units: a signal source, a signal processor, and an application module. Signals pass from the source module to the signal processor, where the signal is transformed to meaningful data which is then passed to the application module that acts on the data. Signal sources can be easily exchanged to allow several amplifier alternatives to be used with little to no modification of the other modules. Among the included prebuilt modules is a signal processing module for classifying P300 signals, and is used along with an included P300 speller application module.

The BCPy2000 framework is a community package enabling BCI2000 modules to be written using the Python scripting language. BCPy2000 modules can be mixed and matched with traditional BCI2000 modules, so that existing modules possessing the desired functionality would not need to be rewritten in Python. In this application, the BCPy2000 framework was used to design the application module, while the

P3SignalProcessing module included with BCI2000 was used to classify P300 responses to stimuli presentation.

### 2.2 Application Design

Included with BCI2000 is a P3Speller application module along with a `P3SignalProcessing` module. The `P3Speller` application employs a parametric matrix, with a 6x6 alphanumeric matrix set as the default. It uses the oddball paradigm with Farwell and Donchin's classical row/column approach for stimulus presentation [12]. Due to the complexity of the design and the exploratory nature of the project, an entirely new BCPy2000 based application module was created rather than modifying the existing `P3Speller` application module. The `P3SignalProcessing` module is used for signal processing, and its output is utilized in the Facebrain application module.

To begin the program, the user navigates to the `BCI2000/batch` directory inside the Facebrain installation folder and executes the `Facebrain.bat` file. `Facebrain.bat` loads the appropriate signal source module (customizable by the user), the `P3SignalSource` module, and the `PythonApplication` module responsible for loading the Python source code present in `BCI2000/python/Facebrain.py`. A fourth unit named the `Operator` module coordinates the communication between those three modules.

Selecting Config brings up a Parameter Configuration panel displaying options for window sizing, matrix sizing, and stimulus/inter-stimulus durations, as well as various other customizable options for the varying components. Selecting Set Config from the `Operator` initializes the application with the parameters set during configuration, and brings up the blank application window. The Start button will become active, and selecting it will show the Home menu.

### 2.3 Menu System

Each menu and submenu of the application consists of an implementation of a P300-style speller. The interface consists of a matrix of selectable options to be chosen by the user via a flashing mechanism. The row/column paradigm has been implemented by default, but the architecture of the application allows for alternate flashing algorithms to be implemented and set via the configuration options if desired. The home menu consists of a 3 x 3 matrix of targets representing nine commonly used Facebook functions. A 7 x 7 matrix is also utilized for spelling words and sentences in the various submenus. Figure 1 illustrates the graphical interfaces for both the home menu and menu option for inputting text, an option similar to other researchers allowing additional freeform text by participants [7].

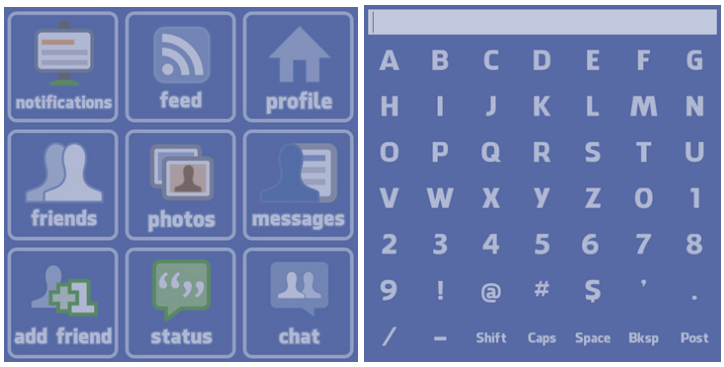



**Fig. 1.** Views of Facebrain's home menu (left) and an example 7x7 speller matrix used for text input (right).

### 2.4 Speller Functionality and Target Classification

The Farwell and Donchin [13] row/column algorithm was chosen for stimulus presentation due to its ease of implementation and longstanding history. For every matrix of targets, the stimuli are grouped into sets of rows and columns, and each set is labeled with a distinct `StimulusCode`. The `StimulusCodes` are then compiled together into a single `Sequence` and randomly shuffled. A dialog is displayed to the user informing them to focus on a target and to count the number of flashes. A countdown is given, and after a short time the rows and columns are flashed in the order present in the sequence. After a sequence has completed, the sequence is randomly shuffled again such that the last stimulus of the completed sequence is not the first element of the new sequence. This continues until the total number of presented sequences equals the `NumberOfSequences` parameter.

At the start of each flash, the `P3SignalProcessing` module is informed of that flash's `StimulusCode` and begins recording an epoch of signal data. The length of the epoch is defined in the `EpochLength` parameter (default: 500ms). The `NumberOfSequences` parameter defines how many epochs are to be averaged before reporting the results to the application module. The reported results contain a log-likelihood ratio that the `StimulusCode` being reported contains the desired P300 response. After all stimuli have been reported, the application module determines the two highest rated `StimulusCodes` and attempts to find an intersection between their corresponding sets of targets. If a single target intersects both sets, it is determined to be the most likely target and is indicated to the user. The system then performs the appropriate function associated with the target, sets appropriate variables, and continues with the program.

## 3 Summary

Facebrain as a P300-based brain-computer interface for accessing Facebook may provide a new level of social inclusion and enhancement to the quality of life for a person with severe motor impairments to the point of being locked-in. This interface incorporates the BCPy2000 framework for the BCI2000 open source package. This interface has been developed and tested in-lab with a signal generator module that mimics eight channels of real EEG but not yet more widely tested with live participants. Although still in its preliminary stages, a working prototype is available and shows promise that the P300 response may be utilized to provide much more than the ability to type. Future iterations may incorporate additional classifiers for asynchronous usage.

## References


1. Vance, A. Facebook: The Making of 1 Billion Users. Bloomberg Businessweek (2012)
2. Nijholt, A., et al., Brain-Computer Interfacing for Intelligent Systems. IEEE Intelligent Systems, 23(3): p. 72-79 (2008)
3. Wolpaw, J.R., et al., Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology, 113(6): p. 767-791 (2002)
4. Randolph, A.B. and M.M. Moore Jackson, Assessing Fit of Nontraditional Assistive Technologies. ACM Transactions on Accessible Computing, 2(4): p. 1-31 (2010)
5. Holz, E.M., et al., Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study. Archives of physical medicine and rehabilitation, 96(3): p. S16-S26 (2015)
6. Moore, M.M., Real-World Applications for Brain-Computer Interface Technology. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2): p. 162-165 (2003)
7. Halder, S., et al., Brain-controlled applications using dynamic P300 speller matrices. Artificial intelligence in medicine, 63(1): p. 7-17 (2015)
8. Kübler, A., et al., A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. Annals Of The New York Academy Of Sciences, 1157: p. 90-100 (2009)
9. Pinegger, A., et al., Control or non-control state: that is the question! An asynchronous visual P300-based BCI approach. Journal of neural engineering, 12(1): p. 014001 (2015)
10. Schalk, G., et al., BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering, 51(6): p. 1034-1043 (2004)
11. Schalk, G. Effective brain-computer interfacing using BCI2000. in IEEE Engineering in Medicine and Biology Society (2009)
12. Aarts, H., B. Verplanken, and A. Van Knippenberg, Predicting behavior from actions in the past: Repeated decision making or a matter of habit. Journal of Applied Social Psychology, 28(15): p. 1355-1374 (1998)
13. Farwell, L.A. and E. Donchin, Talking Off the Top of Your Head: Toward a Mental Prothesis Utilizing Event-Related Brain Potentials. Electroencephalography and Clinical Neurophysiology, 70(6): p. 510-523 (1988)

# Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface

Théophile Demazure[1], Alexander Karran[1], Élise Labonté-LeMoyne[1], Pierre-Majorique Léger[1], Sylvain Sénécal[1], Marc Fredette[1], Gilbert Babin[1]

[1] HEC Montréal, Montréal, QC H3T 2A7, Canada
{theophile.demazure,alexander-john.karran,pierre-majorique.leger,elise.labonte-lemoyne,sylvain.senecal,marc.fredette,gilbert.babin}@hec.ca

**Abstract.** In today's data-driven information technology environment, the ability of humans to sustain attention over long periods of time has become an increasingly important skill. We report work in progress to create a novel passive brain computer interface (pBCI), designed to modulate a user's level of sustained attention in an ecologically valid information system (IS) context. To modulate sustained attention, we take measures of cognitive engagement and vigilance using electroencephalography (EEG) in real time, to form the basis of the BCI, and create a closed neurophysiological feedback loop which adapts elements of a dynamic user interface according to the user's level of sustained attention. The interface utilizes the ERPsim simulation engine to create an ecologically valid IS task supported by a real-life ERP framework.

**Keywords:** Passive Brain-Computer Interface · EEG · Vigilance · Adaptive System · Human-computer interaction · ERP.

## 1 Introduction

Recent advances in information technologies, such as artificial intelligence and robotics are rapidly reshaping the way in which we interact with technology [1]. Tasks commonly performed through human labour are becoming increasingly automated, creating vast subsets of tasks that require a high level of human decision readiness, and a high degree of sustained attention (SA) to monitor complex information systems (IS), and the data they create. Users of modern IS, ranging from critical systems infrastructure, to business logistics, require the ability to quickly synthesize and interpret a wide variety of information, to make correct and timely decisions. However, the rapid adoption of automation for administration and analysis tasks has resulted in a potentially hazardous business mindset that considers the human element as a secondary function [2]. Studies have shown that while automation has increased productivity by reducing information-processing and cognitive load, it has decreased operator decision readiness and on-task safety, and that errors are often the result of a decrease in operator vigilance and SA [3,4], which resulted in a call for research in the IS domain [5].

Previous Research has demonstrated that performance in long duration SA tasks is greatly reduced over prolonged and continuous periods of time [6]. This reduction in SA, termed the vigilance decrement [7] begins to manifest after 20–30 minutes of task engagement, whereupon reaction times and the probability of operator decision errors increase [8]. Thus, the vigilance decrement occurs when signals requiring detection are perceivable to operators, but do not compel changes in the operating environment. Our aim is to create a brain-computer interface that modulates a user's level of SA, by combining measures of task engagement, vigilance and an autoadaptive IS interface which creates attentional signals to encourage changes in the operating environment.

In the following sections of this manuscript we outline the design methodology and process framework for a BCI that monitors, classifies, and modulates a user's ability to maintain a steady state of SA in real time while monitoring a complex logistics task in enterprise system, and report on results from preliminary analysis.

## 2 Artefact Design and Requirement Analysis

Utilising design science methods [9], built upon application strategy 3 of NeuroIS [10] and synthesising previous work in the field of neuroscience concerned with task engagement [11,12,13], we created an iterative design and testing strategy that allowed for rapid development and testing, utilising both synthetic and real neurophysiological data. To capture the requirements of the BCI artefact, we first analysed the needs of an ecologically valid business task, then analysed the needs of an SA task. From this analysis we identified 3 primary requirements:

1. The artefact must represent a real information system (IS) monitoring task.
2. The IS task and its duration must induce and promote a vigilance decrement in the user.
3. The BCI component of the artefact should provide counter measures (CM), to modulate the level of SA, leading to a performance enhancement of its user without obstruction of the IS task.

With regard to the requirement 3, the use of electronic countermeasures within an interface or software artefact, to improve task performance or modulate cognitive workload is currently an area of active research within the human factors community (see NASA HRR [14]).

### 2.1 Implementation

To meet requirements 1 and 2, we created an ecologically valid IS task utilising an enterprise system [15,16] offering the functionality, process, reports, and data to simulate a real-life organisation (i.e. SAP). To produce a vigilance decrement in the participant, we modified an enterprise system simulation called ERPsim (ERPsimLab, Montréal) so that time moves much slower than the speed suggested by the creators [15], extending the duration of the task to 90 minutes. The task itself involves maintaining

stock levels in 3 locations, and participants are asked to make logistical decisions concerning stock allocation. During the simulation participants are required to perform 15 maintenance and 4 decision tasks each of 4 minutes duration. Stock depletion rates are non-uniform and dependent on different demand functions. A maximum stock capacity is provided to force decisions as soon as new stock is received, and all correct, incorrect, and missed decisions are logged for later analysis. Thus, the task was reduced to a monitoring task requiring a high level of sustained attention.

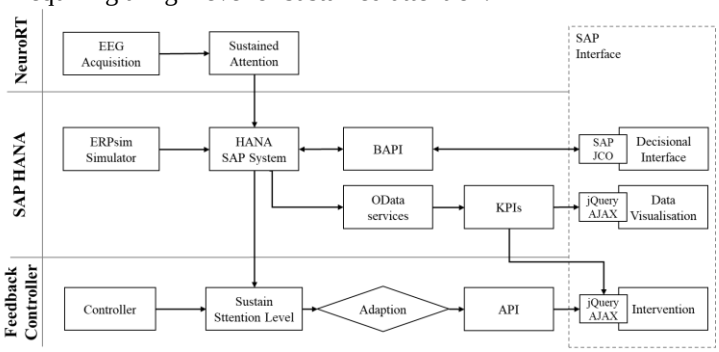



**Fig. 1.** BCI Artefact Software Architecture schematic

The architecture of the BCI artefact has two components, hardware (section 2.2), and software (See Figure 1.), which consists of three elements: 1.) NeuroRT software (Paris, France) implementing real-time data processing to extract SA according to chosen parameters. 2.) SAP HANA (Walldorf, Germany), which deals with back-end operations such as, storing and serving the neurophysiological data, displaying the information dashboard, and running the simulation and storing the data it creates. Odata services allow the creation of the information dashboard which is automatically refreshed using asynchronous AJAX calls, the decision interface is provided using SAP JCO to directly call BAPI. SAP HANA is deployed to support the experiment via a HANA server owned by HEC Montreal. 3.) the Feedback Controller provides the CM mechanism developed in Python, to classify a user's level of sustained attention provided by 1. and served to by 2. All information displayed within the interface follows the concepts of dashboard design [17,18] and is provided through API queries. Furthermore, to modulate the attentional state of the user, interface CM are applied using a dynamic colour palette to indicate to the user their level of attention, such that the screen background is: white= high, amber = below optimum, red = poor. Thus, during active task phases, if the user remains in a heightened state of SA, there are no CM and the interface remains unchanged.

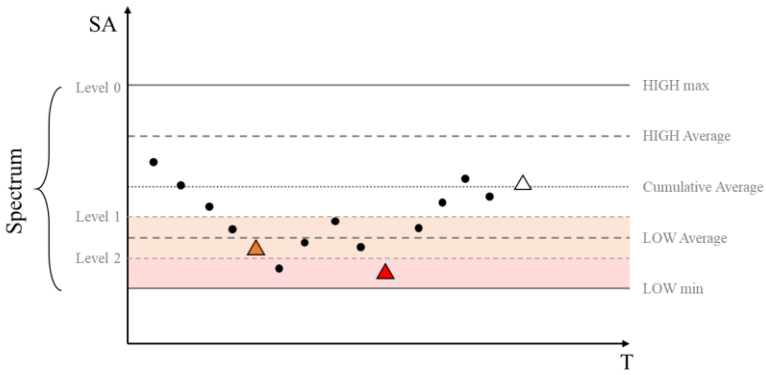



**Fig. 2.** User-specific attention spectrum

## 2.2 Neurophysiological Methods

We utilise Pope et. al's [11] engagement index to provide real-time assessment of a user's attentional state and to drive the neurofeedback mechanism of the simulation task interface. This index has been previously used to observe a vigilance decrement [12,13], making it an ideal candidate measure for this BCI. Following [12] we used a sensor hardware platform consisting of a 32 electrode EEG (Brainvision, Morrisville, NC), to measure variations in brainwave activity in the θ (4-7Hz), α (8-12Hz) and β (13-21Hz) bands from F3, F4, O1, O2 on the international 10–20 system [19]. The SA index is calculated using β (power) divided by α (power) plus θ (power).

From a proposed participant pool of N= 30, 12 participants (6 female) aged 18-43 (Avg. = 24.89), have so far taken part in the study. Participants were of good health and possessed normal or corrected to normal vision, all signed consent in line with the University's ethics board and compensated 50 CAD. Participants were provided with a mouse and keyboard and sat approximately 80 cm from a 24" computer screen.

The experimental task is split into 2 parts: calibration, lasting 22 mins, and testing, lasting 90 mins. During calibration the SA index is personalised to individual participants to create a spectrum scale of values ranging from high to low. Calibration is composed of a 1 min baseline (passive observation), then an engagement task of 10 mins, then a 1 min baseline and a vigilance task of 10 mins. We then compute thresholds for the individual that allow a variable spectrum of index values for user SA state, that fluctuates during part 2 of the experiment (see Figure 2.) in response to changes in mean SA threshold levels over time. These thresholds form the basis of a set of rules that are the coded to create a "fuzzy logic" classifier that classifies SA into three levels where 0 = high SA, 1 = moderate SA and 2 = low SA. These classifications are then used by the feedback controller to produce CM (i.e. changes in the background colour) to modulate SA in the user.

## 3 Evaluation

Initial development of the BCI was completed iteratively using a combination of simulated EEG data, observation, and hands-on tests. The ERP simulation and task interface were built and tested, with and without adaptations, using test participants to provide feedback concerning the task, the simulation, and the overall experience.

The full experimental procedure (calibration and test) is run to evaluate the current BCI - IS artefact. Participants are randomly assigned to 1 of 3 conditions: no CM; continuous CM; event-based CM. During continuous CM, modulation occurs during the whole experiment. Event based CM consists of modulation only during event phases. User performance during the simulation is measured through actions per minute (APM), percentage of decision errors (PDE) and simulation score, this is then compared with SA level to assess the impact of SA modulations. Participants complete a questionnaire at the end of the procedure to provide a subjective assessment of perceived workload, level of boredom and affective response towards the task and interface. The questionnaire is composed of the raw-TLX [20], a shorter version of the NASA-TLX [21], the Boredom Experience Scale [22,23], and a SAM Scale [24].

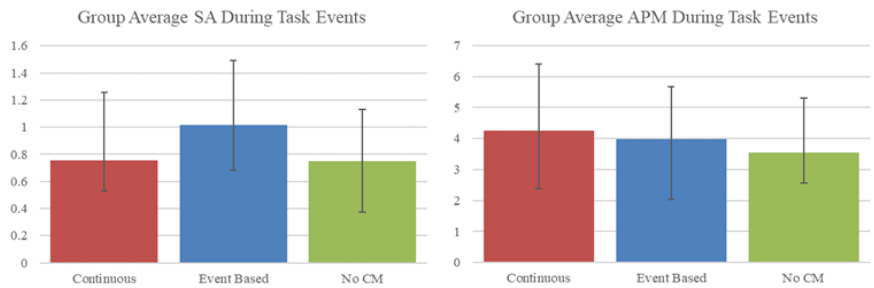



**Fig. 3.** Preliminary results n=12, n=4 per group showing mean SA and APM across conditions during task events

Figure 3 displays the mean SA and mean APM values for each group (n=4) during task events for the currently available data, error bars represent variance within group. In this figure we observe almost no discernable mean difference in SA between the no CM (μ 0.75) and continuous CM (μ 0.76) groups, however given the within group variance, this may not reflect the overall strength of the effect once all data is collected. Looking at the APM difference between the same groups, we see that the trend indicates that those in the continuous CM group perform more task actions within active task events. However, if this effect correlates with less decision errors remains to be determined. Looking at the effect on SA and APM for the event-based CM group, shows a higher mean SA (μ 1.01) than both no CM, and continuous CM groups and a higher APM than the no CM group. This potentially indicates that event-based CM promotes higher SA, but not necessarily more actions within active task events, when compared to the continuous CM group, whether the effect of increased SA and APM equates to less decision errors remains to be determined after data collection is complete.

## 4 Work in progress and next steps

The data so far indicate that the BCI artefact has a positive modulating effect on a user's level of SA, and positively influences the actions they take during active event periods during the enterprise system simulation. However, it is yet to be determined if this effect remains positive across both CM groups after data collection and analysis is complete. Furthermore, the relationship between modulated SA, increased APM and decision errors remains to be explored, as does the correlation between a qualitative assessment of performance, workload, and SA with quantitative observations.

In addition to utilising EEG to measure SA for this project, we also employed functional Near Infrared Spectroscopy (fNIRS) concurrently to measure changes in SA. fNIRS measures the haemodynamic response function of cortical areas to infer synaptic network activation. From this we seek to gain a deeper understanding of SA and apply machine learning to these data, to determine if fNIRS could also be used classify SA in real-time. Furthermore, we seek to utilise both fNIRS and more sophisticated derivatives of EEG data to create a hybrid BCI with the potential to provide a more granular, dynamic, interface environment and more robust SA assessment. We look forward to reporting our results once data collection and analysis is complete.

## References

1.Autor, David H. (2015). "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." Journal of Economic Perspectives, 29(3): 3–30.
2.Warm, J.S., Matthews, G., & Finomore, V.S. (2008) Workload and stress in sustained attention. In P.A. Hancock and J.L. Szalma (Eds.), Performance under stress, pp.115-141. Aldershot, UK: Ashgate Publishing.
3.Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. Human Factors, 38, 311–322.
4.De Boer, R.; Dekker, S. (2017). Models of Automation Surprise: Results of a Field Survey in Aviation. *Safety*, *3*, 20.
5.Brocke, J. V., Riedl, R., & Léger, P. M. (2013). Application strategies for neuroscience in information systems design science research. Journal of Computer Information Systems, 53(3), 1-13.
6.Mackworth, N. H. (1948). "The breakdown of vigilance during prolonged visual search." Quarterly Journal of Experimental Psychology 1(1): 6–21.
7.Parasuraman, R., Warm, J. S., & See, J. E. (1998). Brain systems of vigilance.
8.Grier, R. A., et al. (2003). "The vigilance decrement reflects limitations in effortful attention, not mindlessness." Human Factors 45(3): 349–359.
9.Von Alan, R. H., et al. (2004). "Design science in information systems research." MIS quarterly 28(1): 75–105.
10.Brocke, J. V., Riedl, R., & Léger, P. M. (2013). Application strategies for neuroscience in information systems design science research. Journal of Computer Information Systems, 53(3), 1-13.
11.Pope, A. T. B., Edward H.; Bartolonne, Debbie S. (1995). "Biocybernetic system evaluates indices of operator engagement in automated task."

12.Mikulka, P. J., et al. (2002). "Effects of a Biocybernetic System on Vigilance Performance." Human Factors: The Journal of the Human Factors and Ergonomics Society 44(4): 654–664.
13.Léger, P. M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. Computers in Human Behavior, 34, 273-283
14 NASA HRR "Development of a Software and User Interface to Support Scenario Modeling of Astronaut Schedules to Aid in the Selection of Fatigue Countermeasures within the Behavioral Health and Performance Dashboard (BHP-DS)"
15.Léger, P.-M. (2006). "Using a simulation game approach to teach enterprise resource planning concepts." Journal of Information Systems Education 17(4): 441.
16.Léger, P.-M., Robert, J., Babin, G., Pellerin, R. and Wagner, B. (2007), ERPsim, ERPsim Lab, HEC Montréal, Montréal, Qc.
17.Few, S. (2006). "Information dashboard design."
18.Few, S. (2012). Show me the numbers: Designing tables and graphs to enlighten, Analytics Press.
19.Jasper, H. H. (1958). "Report of the committee on methods of clinical examination in electroencephalography." Clinical Neurophysiology 10(2): 370-375.
20.Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting, Sage Publications Sage CA: Los Angeles, CA.
21.Hart, S. G. and L. E. Staveland (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, Elsevier. 52: 139–183.
22.van Tilburg, W. A. and E. R. Igou (2012). "On boredom: Lack of challenge and meaning as distinct boredom experiences." Motivation and Emotion 36(2): 181–194.
23.Vodanovich, S. J. and J. D. Watt (2016). "Self-report measures of boredom: An updated review of the literature." The Journal of Psychology 150(2): 196–228.
24.Bradley, M. M. and P. J. Lang (1994). "Measuring emotion: the self-assessment manikin and the semantic differential." Journal of behavior therapy and experimental psychiatry 25(1): 49–59.

# Towards Designing Robo-Advisors for Unexperienced Investors with Experience Sampling of Time-Series Data

Florian Glaser[1], Zwetelina Iliewa[2], Dominik Jung[1,3], and Martin Weber[4]

[1]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[2]Max Planck Institute for Research on Collective Goods, Bonn, Germany
[3]Karlsruhe Decision and Design Lab (KD2Lab), Karlsruhe, Germany
[4]University of Mannheim, Center for European Economic Research (ZEW), Mannheim, Germany
{florian.glaser, d.jung}@kit.edu, iliewa@coll.mpg.de, weber@bank.bwl.uni-mannheim.de

**Abstract.** We propose an experimental study to examine how to optimally design a robo-advisor for the purposes of financial risk taking. Specifically, we focus on robo-advisors which are able to (i) ”speak” the language of the investors by communicating information on the statistical properties of risky assets in an intuitive way, (ii) ”listen” to the investor by monitoring her emotional reactions and (iii) do both. The objectives of our study are twofold. First, we aim to understand how robo-advisors affect financial risk taking and the revisiting of investment decisions. Second, we aim to identify who is most affected by robo-advice.

**Keywords:** Robo-Advisory · Financial Risk Taking · Emotion Regulation · Biofeedback · Physiological Arousal

## 1 Introduction

Without robo-advisors, taking financial risks is a challenging task for retail investors. Consequently, a new form of financial advisory, so called robo-advisors have become more and more popular in retail and private banking [1], due to their ability to provide unbiased financial advice for retail investors at low cost [2].

Robo-advisors are digital platforms using information technology to guide customers through an automated (investment) advisory process based on interactive and intelligent user assistance components [1, 3]. In particular, they can be helpful tools for investors with low financial knowledge, as well as investors who are susceptible to making financial mistakes [1].

Hence, robo-advisors might have the potential to be useful investment aids to guide people to a share of risky investments which fits their needs and their preferences [3]. Unexperienced investors are not only the largest part of the population; they are also subject to numerous known fallacies in financial decision-making. Across countries,

households shy away from risky financial assets [4] and those who take financial risks, often make costly financial mistakes (see, e.g., [5]). In Germany, only 14% of the population own shares [6], a number which is considerably lower than in the UK or US. Consequently, there is considerable research interest in financial advisory because these investors need good decision support.

However, the design of such systems is not an easy matter, and we argue that the sources of financial mistakes vary for different individuals (e.g. low statistical skills) and in different situations (e.g. stress, cognitive load etc.). Consequently, computer-based decision support systems like robo-advisors should take into account (i) individual user characteristics and (ii) situational factors such as the user's internal state. Previous research in information systems has illustrated the importance of considering individual user characteristics when designing decision support systems. The importance of leveraging user models to individualize decisional guidance and explanations has been recognized many years ago [7, 8]. Much work has been done in this field and the stream of research on designing "knowledge interfaces" is still evolving [9], providing new functions, for example, gamification features [10], advanced presentation (virtual and augmented reality) as well as intelligent adaptation leveraging user models on the basis of multi-modal user monitoring. Multi-modal user monitoring includes measurement of users' physiological states (e.g., arousal or cognitive load) which help to assess the user's internal states.

In conclusion, recent information systems research has recognized user characteristics as an important factor for adoption and use of different IS types, for example, in collaboration systems [11], group-decision-making [12], or multi-channel financial services [13, 14]. Users' internal states play a crucial role in information processing and decision behaviour [15–17]. Thus, to successfully support computer-supported investment decisions, we consider a deeper understanding of individual characteristics and internal states as an important research gap.

The present paper presents an experimental study to examine how to optimally design a robo-advisor for the purposes of financial risk taking, thus targeting this research gap. On the basis of a short literature review, we present our experimental design. The work concludes with a summary and an outlook.

## 2 Designing Financial Decision Support Systems

### 2.1 Financial Decision Support Systems

Some studies have attempted to design financial information systems that, for instance, restrict reinvestment or provide decisional guidance [18]. Another stream of research is taking a closer look at how well individuals actually understand the decisions they are asked to make (e.g. [19]). These studies show that individuals generally find it difficult to process risk and probabilities (see e.g. [20]) which often leads to a skewed view of the risk or volatility associated with an investment alternative. By altering the way information on the underlying statistical distribution is presented

from a mere description towards experiencing the distribution through simulation (i.e. experiential presentation), individuals obtain a better understanding of the investment alternatives and are more inclined to make a risky investment [21, 22].

### 2.2 Robo-Advisory

Initial research in the area of robo-advisory has, for instance, focused on the design of the risk election and modelling of robo-advisors [23]. Tertilt and Scholz reviewed the profiling steps of existing robo-advisory solutions, investigating the risk profiling. Jung et al. derived different design principles based on literature review and expert interviews, and evaluated them in a design science approach [2]. Other work has focused on the legal limitations of robo-advisory [24], or comparing the cost and quality of robo-advisors and human advisory [25].

Nevertheless, only a few studies have addressed the design of robo-advisor user interfaces so far [3]. We argue that there is an increasing need to target this shortcoming and focus on the design of robo-advisory platforms. For instance, service and user interface design, customer behavior, and risk measurement and modelling have been identified as areas with a pressing need for investigation [3].

## 3 Proposed Experimental Study

In this initial project we plan to design and implement a prototypical robo-advisor, which provides users with basic and advanced experientially presented risk information. In the laboratory experiment, we will compare how well different types of individuals understand risk, what their investment decisions are, and how satisfied and confident they are depending on whether they use this new system or a system that provides the same information in a simpler way. Prior to the experimental study, we will conduct a small pilot test with members of both groups to elicit additional presentation design requirements and suggestions.

The experiment will be a repeated between-subject design with two groups of participants with low and high levels of expertise in statistics, respectively. Participants in both groups are randomly assigned to one of the two treatment conditions, basic and advanced experiential risk information presentation. In particular, we use a 2×2 between-subject design to test for the impact of two potentially important characteristics of the design of robo-advisors, separately and in combination, against the benchmark of no advice. Subjects are randomly distributed in the four treatments. The first characteristic of the design of a robo-advisor is connected to its ability to ”speak” the language of the investors. It consists in communicating the information about the probability distribution of the risky asset by allowing the investors to sample from it instead of showing them a formal description. The second characteristic is related to the ability of the robo-advisor to ”listen” to the investor. It consists of continuously monitoring the emotional reaction of the investors over all stages of the decision making process and training the investors how to regulate their emotions. The experiment will be conducted with 200 subjects and divided into four treatments of 50 subjects

respectively. A pretest will be conducted prior to the experiment for the purposes of calibrating the duration of filler tasks and testing for sufficient cross-sectional variability of our measures of personal characteristics. The experimental procedure is illustrated in Figure 1.

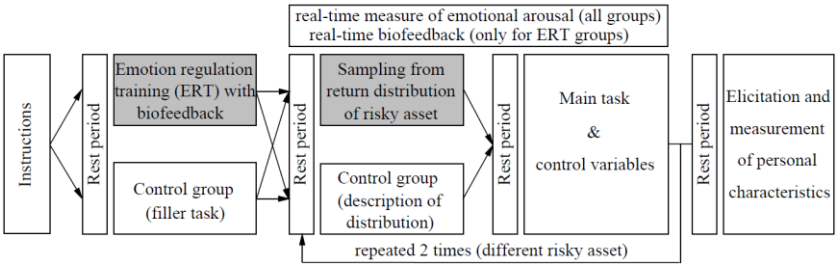



**Fig. 1.** Proposed design of the initial experimental study

To gain insight on why robo-advice might potentially affect financial risk taking, we measure several covariates of the decision-making process. On the asset level, we obtain subjects' understanding of the probability distribution and their risk perception. For every decision, we elicit subjects' confidence and decision time. For every stage of the decision-making process (information acquisition, decision, realization of outcome), we measure subjects' heart rates (physiological arousal) with the aim of identifying changes in their internal states and relating them to individual characteristics and treatments. For that purpose, we will rely on the experimental platform Brownie, following the Brownie guideline for experimental research in NeuroIS [26]. Participants fill in a questionnaire about their statistics expertise (possibly using an objective measure as the Berlin numeracy task), investment experience, demographic information, risk aversion etc. They are then presented with the treatment (risk information presentation) and subsequently given the task to judge several investment alternatives with respect to their risk / volatility and to indicate which investment alternative they would choose. The final questionnaire surveys decision confidence, satisfaction, etc.

We expect that our experiments help explain how different individuals need to be presented with risk information such that they understand it better and, as a result, are able to make better investment decisions. Our research also helps to develop recommendations for designing better financial information systems for retail investors.

## 4 Expected Results and Conclusion

Decision support systems like robo-advisors have the potential to be useful investment aids and more research should be devoted to examine how to design them. This experimental study is intended to build the basis for future research at the intersection of Finance and Information Systems. Intelligent systems that adapt to individual users to provide advanced support for investment decisions are considered as an important research domain with a strong impact on society.

In particular, we are interested in examining the temporal stability of learning and the effect of stress. The quality of and capacity for information processing decreases with higher cognitive load and at higher stress levels. For instance, high investments likely lead to higher stress levels for individuals, and that could argue for the need to present information differently (e.g. design a stress-sensitive adaptive system). In future research, we plan to add one treatment variable (stress) and the corresponding treatments. With regard to the temporal stability of learning, we plan to conduct repeated measures studies, for example by inviting participants to multiple experimental sessions in which they are repeatedly given the task to judge several investment alternatives.

## References

## References


1. Sironi P (2016) FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification. John Wiley & Sons
2. Jung D, Dorner V, Weinhardt C et al. (2017) Designing a Robo-Advisor for Risk-Averse, Low-Budget Investors. Electronic Markets, Special Issue on FinTech and the transformation of the Financial Industry
3. Jung D, Dorner V, Glaser F et al. (2018) Robo-Advisory: Digitalization and Automation of Financial Advisory. Business & Information Systems Engineering 60(1): 81–86
4. Badarinza C, Campbell JY, Ramadorai T (2016) International comparative household finance. Annual Review of Economics 8: 111–144
5. Calvet LE, Campbell JY, Sodini P (2007) Down or out: Assessing the welfare costs of household investment mistakes. Journal of Political Economy 115(5): 707–747
6. Deutsches Aktieninstitut (2015) Aktionärszahlen des Deutschen Aktieninstituts. https://www.dai.de/files/dai_usercontent/dokumente/studien/2016-02-09%20DAI%20Aktionaerszahlen%202015%20Web.pdf
7. Gregor S, Benbasat I (1999) Explanations from intelligent systems: Theoretical foundations and implications for practice. MIS quarterly: 497–530
8. Silver MS (1991) Decisional guidance for computer-based decision support. MIS quarterly: 105–122
9. Gregor S, Maedche A, Morana S et al. (eds) (2016) Designing knowledge interface systems: Past, present, and future. DESRIST 2016
10. Feil S, Kretzer M, Werder K et al. (eds) (2016) Using gamification to tackle the cold-start problem in recommender systems. ACM
11. Schacht S, Morana S, Urbach N et al. (2015) Are you a Maverick? Towards a Segmentation of Collaboration Technology Users
12. Knierim MT, Jung D, Dorner V et al. (eds) (2017) Designing Live Biofeedback for Groups to Support Emotion Management in Digital Collaboration. Springer
13. Hummel D, Schacht S, Maedche A (2016) Determinants of Multi-Channel Behavior: Exploring Avenues for Future Research in the Services Industry
14. Hummel D, Schacht S, Maedche A (2017) Designing Adaptive Nudges for Multi-Channel Choices of Digital Services: A Laboratory Experiment Design
15. Astor PJ, Adam MTP, Jerčić P et al. (2013) Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. Journal of management information systems 30(3): 247–278
16. Dimoka A, Davis FD, Gupta A et al. (2012) On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS quarterly: 679–702
17. Riedl R, Davis FD, Hevner AR (2014) Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. Journal of the Association for Information Systems 15(10): I
18. Looney CA, Hardin AM (2009) Decision support for retirement portfolio management: overcoming myopic loss aversion via technology design. Management Science 55(10): 1688–1703

19. Ehm C, Kaufmann C, Weber M (2014) Volatility inadaptability: Investors care about risk, but cannot cope with volatility. Review of Finance 18(4): 1387–1423
20. Jung D, Dorner V Decision Inertia and Arousal: Using NeuroIS to Analyze Bio-Physiological Correlates of Decision Inertia in a Dual-Choice Paradigm. In: Information Systems and Neuroscience 2017, pp 159–166
21. Bradbury MAS, Hens T, Zeisberger S (2014) Improving investment decisions with simulated experience. Review of Finance 19(3): 1019–1052
22. Kaufmann C, Weber M, Haisley E (2013) The role of experience sampling and graphical displays on one's investment risk appetite. Management Science 59(2): 323–340
23. Tertilt M, Scholz P (2017) To Advice, or Not to Advice—How Robo-Advisors Evaluate the Risk Preferences of Private Investors
24. Fein ML (2015) Robo-Advisors: A Closer Look. Available at SSRN 2658701
25. Fisch JE, Turner JA, Center PP (2017) Robo Advisers vs. Humans: Which Make the Better Financial Advisers?
26. Jung D, Adam M, Dorner V et al. (2017) A practical guide for human lab experiments in information systems research: A tutorial with Brownie. Journal of Systems and Information Technology 19(3/4): 228–256

# Neural Correlates of Human Decision Making In Recommendation Systems: A Research Proposal

Naveen Zehra Quazilbash, Zaheeruddin Asif

Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan
{nquazilbash, zasif}@iba.edu.pk

**Abstract.** Significant research has been conducted on human decision making behavior in recommendation systems during the past decade, yet it remains a challenge to design effective and efficient recommendation systems so that they not only produce useful suggestions and ease the decision making task but also turn it into a pleasurable experience. Algorithms have been designed based on research that highlight individual theoretical constructs yet there is an absence of a comprehensive model of human decision-making. This research offers an insight into the core of this issue by examining the neural correlates of human decision-making using Electroencephalography (EEG). The insights generated maybe used to construct a comprehensive model of human decision making in recommendation systems and generate new design principles for the same.

**Keywords:** Decision making · Recommendation systems · EEG · Source localization

## 1 INTRODUCTION

Picking an option out of a set of recommendations is a crucial skill for users aiming to solve real-life decision-making problems. Which book to read, which movie to watch, who to be friends with, which gadget to buy and what items to shop for grocery are a few examples where people may choose to go for a particular suggestion. While taking complex decisions, people usually rely on advice provided by others [1]. Whereas, in online settings, Recommendation Systems (RS) do the job [2, 3]. The software that facilitates human decision making by providing useful suggestions is called a recommendation system.

In the context of recommendation systems, the process of arriving at a decision is a two-way collaborative process. It is collaborative in a way that the recommender suggests something and the chooser decides or may give some feedback on it, based on which the recommender may improve the suggestion, thereby helping the chooser reach a decision. Hence, it is necessary to maintain this collaboration between the user and recommendation system [4, 5, 6, 7]. A better understanding of how people decide or make choices may help in achieving this goal.

Since the process of choosing or decision making is tiring and effort-intensive [8], a good choice would refer to the one that involves little effort and time [7]. Therefore, it

would be useful to determine the factors influencing the choice behavior, because this information can be used for a reduction in the effort applied in decision making and an improvement in generating accurate suggestions [9].

Previous research has shown that there are four major categories of brain processes, namely, social, emotion, cognitive and decision making processes; each involving specific brain areas. Neurological studies have shown that a combination of brain areas get activated when a certain process is being performed. The specific brain areas map onto specific constructs such as ambiguity, risk, emotion and cognitive calculations [10]. However, there is no one-to-one mapping between these processes and the brain areas that get activated.

Since, there does not exist a one-to-one mapping between the brain areas and processes i-e only one factor contributing towards any particular process, say decision making, it is pertinent to explore the network of factors that contribute towards human decision making process in RSs. This network, referred to as "model" in the proposed study, will help in opening up the black box of human brain processes, which in turn may provide useful insights for improving the design of recommendation systems. This conforms to the call by Taylor to design effective ways of refining information systems by modelling internal neurological actions [11].

Developing a better understanding of human decision making in recommendation systems by using data collected through surveys via questionnaires and interviews is a difficult task. Most of the studies exploring constructs like emotion and trust possess self-reporting bias, owing to their behavioral nature, a case in point being study conducted by Hu and Pu [12]. Self-reported data using questionnaires is limited to conscious perceptions and thoughts, whereas in real life information processing and unconscious perceptions impact human behavior [13]. Due to this influence it is hard to develop an accurate understanding of the IT related human behavior using only the self-reported data.

In order to address the challenge of self-reporting bias, researchers are now using neurophysiological tools to measure brain signals directly instead of asking the subjects [10, 14, 15]. Functional brain imaging has proven to be a promising area in explicating the unanswered questions in fields of psychology, marketing and economics and recently has found its way into Information Systems as well [10, 15, 16]. In the proposed study, Electroencephalography (EEG) is used to capture electrical brain signals, which are then used for localization of their source. This brain source will be mapped onto an IS construct which reflects the participation of this construct in the decision making process.

Combining the knowledge of HDM from social sciences and tools from cognitive neuroscience this study aims to develop a model of constructs that pertain to human decision making in recommendation systems thereby answering the following research questions:

- What are the individual neural correlates of human decision making in content-based book recommendation systems?

- Which dimensions of HDM manifest themselves simultaneously when a decision making activity is performed in content-based book RSs?

In order to address afore mentioned research questions, we intend to conduct a study comprised of complementary empirical and behavioral parts. In this way, we plan to explore the neural correlates of human decision making in recommendation systems. It would help researchers understand how recommendation agents influence individual cognition during multiple choice decision making.

The remaining paper is structured as follows. The following section presents and discusses the proposed hypotheses. Research methodology is discussed in section 3. Section 4 presents a discussion on subjective justification of methodology along with expected contributions in various directions.

## 2 RESEARCH MODEL

Since there does not exist a one-to-one mapping of brain areas contributing toward a particular brain process, instead multiple brain areas get activated when some activity is performed by a human subject, therefore, we expect that decision making activity in content-based book recommendation systems would span simultaneously over different brain areas.

*H1 decision making activity in content-based book recommendation systems span simultaneously over different brain areas.*

*H2 decision making activity elicits more cerebral activity in some brain areas than the others.*

Since any activity excites multiple brain areas simultaneously, so we propose that multiple IS constructs play their role in it. These constructs are referred to as dimensions of human decision making in RSs. Prior research has shown that these constructs tend to include risk, uncertainty, and ambiguity, theory of mind, calculation, distrust, risk and emotion.

The basis of social adaptive learning, a categorical way of following advice is fundamentally rewarding [5] i-e the brain process and/or IS construct behind advice following is that of reward. The experimental fMRI study of human subjects shows that greater positive BOLD responses were generated in reward sensitive brain areas when recommendations were given to support decision making in comparison to decision making that was not supported by recommendations. Two outcomes can be derived from these results that could benefit the design of our study. Firstly, it can be deduced from this study that choice decision making in recommendation systems is a type of social learning activity. Secondly, it links **reward** construct with the choice decision making in advice taking.

*H3a Reward is a dimension of decision making in recommendation systems*

People rely on advice from others for reaching good choices therefore role of recommendation systems providing suggestions to the user can be termed as a social learning activity [5]. This interaction between the user and the recommendation system reflects the approach of deducing how the others are thinking and further predicting their behavior. This theoretical concept is referred to as "*theory of mind*" [17]. The choice activity in recommendation systems is not only social but also calculative in nature therefore Bhatt and Camerer [18] proposed an integration of these two components of "**theory of mind**" by identifying a distinctive neural correlate for each of them.

*H3b Theory of Mind is a dimension of decision making in recommendation systems*

One of the most widely explored constructs in NeuroIS research is Trust. There had been several studies that identify the role of trust in disparate information systems and these studies contribute some fine findings related to the influence of trust measure in different scenarios. (For example (René Riedl et al. 2014)'s [19] study on trust in anthropomorphic decision aids and (Dimoka 2010)'s [20] exploration of trust and distrust constructs). Familiarity with RS builds users' trust; users' beliefs about the degree to which the RSs understand them and are personalized for them are key factors in RS adoption [21]. Whereas on the other hand a lesser emphasis has been laid on exploring the dynamics of distrust. While inferring how other people will think, brain areas pertaining to distrust may also get activated [22, 17].

*H3c Distrust is a dimension of decision making in recommendation systems.*

# 3 PROPOSED METHODOLOGY

This proposal makes use of Electroencephalography (EEG) (a neurophysiological tool) Event Related Potential (ERP), to capture brain signals of human subjects. It is to be done by presenting a stimulus multiple times and then averaging the output signal. This averages out the noise signal and leaves only the signal that was generated in response to the stimulus. Afterwards, sLoreta [23, 24] is used for localization of sources of the signals. These sources can be mapped onto constructs that contribute toward decision making activity. After the empirical part the subjects are supposed to undergo a complementary behavioral part of the study to enhance ecological and external validity of the study. In the behavioral part the participants will be required to fill out a questionnaire related to the activity.

## 3.1 Participants

We plan to recruit thirty participants for the study by posting around the campus. Right-handed healthy individuals with no reported history of any sort of brain injury and perfect and/or corrected-to-perfect eyesight will be recruited for the experiment. These subjects will be aged between 20 and 45 years old book seekers. The study will

be reviewed by Board of Advanced Studies and Research (BASR) before commencement. This sample demographic is based on the analysis of the book-crossing[1] dataset.

### 3.2 Experimental Procedure

Every participant will be tested individually and will be briefed about the study beforehand. The instructions will be shared with the participant in written as well. Subjects will be asked about their favorite book through a subject information form before starting the experiment. The subject will be shown a fixation cross for 10 seconds. The content-based recommendations from an online customized book store will then be showed to the subject. The EEG signals will be recorded starting from the time the fixation cross is shown. The recommendations will be shown as visual stimulus for 15 seconds. Altogether each subject will go through 5 runs of activity with a wash-out time of 5 minutes in between runs.

### 3.3 EEG Acquisition and Analysis

The electroencephalography (EEG) signals will be acquired via MDX Neuro-Pro 32 equipment. 19 channels will be used and the data will be acquired at 200 Hz/s of sampling rate. The electrodes will be positioned as per international 10-20 system. The acquired data will be preprocessed by average referencing.

Finally, source localization is performed by comparing fixation cross and recommendation files in Loreta [23, 24]. This will reveal the brain areas (Brodmann Areas[2]) activated during the decision making activity. In particular, we plan to employ sLoreta [23, 24] for this purpose as it identifies the source with the strongest signal activation.

## 4 DISCUSSION AND EXPECTED CONTRIBUTIONS

Human decision making in recommendation systems is a debated issue in literature. However, little attention has been diverted toward uncovering the neural underpinnings of such activity. In this proposal we suggest that development of a model of neural correlates of human decision making in recommendation systems will contribute towards better design of RSs. Despite the development of research in this area, our understanding would remain constrained if we do not explore the role of the constructs involved in this activity.

Our objective is to develop this model by complementing NeuroIS data with behavioral data to not only to avoid self-reporting bias but also ensure ecological validity via

---

[1] The Book Crossing dataset was collected by Cai-Nicolas Ziegler in 2004. It contains 278,858 users' anonymized demographic data about books.

[2] German scientist Korbinian Brodmann named different regions of the brain based on the cyto-architectural structure of neurons. These areas are referred to as Brodmann Areas.

triangulation of measures. Our study is likely to deliver insights leading to better comprehension of the psychological underpinnings of recommendation influenced decision making. Another expected contribution of this study is to provide insights into implementation of embedded systems since the world is moving towards mind controlled gadgets and robots and decision is one of the normal human conduct that will be shown by the robots in future. Finally, we expect that our work will contribute towards progressing NeuroIS research by using neurophysiological tools to report human behavior in IS thereby help in opening the black box of human brain and thus in designing better systems with improved perceived usefulness.

**Acknowledgements.** We are grateful to the BCI Lab at Department of Physics, University of Karachi for helpful suggestions in drafting this paper and providing facilities.

## References

1. Bonaccio S, Dalal RS. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. Organizational Behavior and Human Decision Processes. 2006 Nov 1;101(2):127-51.
2. Xiao B, Benbasat I. E-commerce product recommendation agents: use, characteristics, and impact. MIS quarterly. 2007 Mar 1;31(1):137-209.
3. Xiao B, Benbasat I. Research on the use, characteristics, and impact of e-commerce product recommendation agents: A review and update for 2007–2012. InHandbook of Strategic e-Business Management 2014 (pp. 403-431). Springer Berlin Heidelberg.
4. Chen L, de Gemmis M, Felfernig A, Lops P, Ricci F, Semeraro G. Human decision making and recommender systems. ACM Transactions on Interactive Intelligent Systems (TiiS). 2013 Oct 1;3(3):17.
5. Biele G, Rieskamp J, Krugel LK, Heekeren HR. The neural basis of following advice. PLoS biology. 2011 Jun 21;9(6):e1001089.
6. Jannach D, Hegelich K. A case study on the effectiveness of recommendations in the mobile internet. InProceedings of the third ACM conference on Recommender systems 2009 Oct 23 (pp. 205-208). ACM.
7. Jameson A, Willemsen MC, Felfernig A, de Gemmis M, Lops P, Semeraro G, Chen L. Human decision making and recommender systems. InRecommender Systems Handbook 2015 (pp. 611-648). Springer, Boston, MA.
8. Cerf M, Garcia-Garcia M, Kotler P. Consumer Neuroscience. MIT Press; 2017 Nov 9.
9. Jannach D, Zanker M, Ge M, Gröning M. Recommender systems in computer science and information systems–a landscape of research. InInternational Conference on Electronic Commerce and Web Technologies 2012 Sep 4 (pp. 76-87). Springer, Berlin, Heidelberg.
10. Dimoka A, Pavlou PA, Davis FD. Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. Information Systems Research. 2011 Dec;22(4):687-702.
11. Taylor JG. Future Directions for Neural Networks and Intelligent Systems from the Brain Imaging Research. InFuture Directions for Intelligent Systems and Information Sciences 2000 (pp. 191-212). Physica, Heidelberg.

12. Hu R, Pu P. A comparative user study on rating vs. personality quiz based preference elicitation methods. InProceedings of the 14th international conference on Intelligent user interfaces 2009 Feb 8 (pp. 367-372). ACM.
13. Lieberman MD. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol.. 2007 Jan 10;58:259-89.
14. Dimoka A, Davis FD, Gupta A, Pavlou PA, Banker RD, Dennis AR, Ischebeck A, Müller-Putz G, Benbasat I, Gefen D, Kenning PH. On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS quarterly. 2012 Sep 1:679-702.
15. Brocke JV, Riedl R, Léger PM. Application strategies for neuroscience in information systems design science research. Journal of Computer Information Systems. 2013 Mar 1;53(3):1-3.
16. Riedl R, Léger PM. Fundamentals of NeuroIS. Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg. 2016.
17. Pavlou P, Davis F, Dimoka A. Neuro IS: the potential of cognitive neuroscience for information systems research. ICIS 2007 Proceedings. 2007 Dec 31:122.
18. Bhatt M, Camerer CF. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. Games and economic Behavior. 2005 Aug 1;52(2):424-59.
19. Riedl R, Mohr PN, Kenning PH, Davis FD, Heekeren HR. Trusting humans and avatars: A brain imaging study based on evolution theory. Journal of Management Information Systems. 2014 Apr 1;30(4):83-114.
20. Dimoka A. What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. Mis Quarterly. 2010 Jun 1:373-96.
21. Komiak SY, Benbasat I. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS quarterly. 2006 Dec 1:941-60.
22. Winston JS, Strange BA, O'Doherty J, Dolan RJ. Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nature neuroscience. 2002 Mar;5(3):277.
23. Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol. 2002 Jan 1;24(Suppl D):5-12.
24. Pascual-Marqui RD. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv preprint arXiv:0710.3341. 2007 Oct 17